THE BIOLOGY AND EVOLUTION OF THE MOLLICUTES/MYCOPLASMA-RELATED ENDOBACTERIA OF ARBUSCULAR MYCORRHIZAL FUNGI

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mizue Naito

August 2014

THE BIOLOGY AND EVOLUTION OF THE MOLLICUTES/MYCOPLASMA-RELATED ENDOBACTERIA OF ARBUSCULAR MYCORRHIZAL FUNGI

Mizue Naito, Ph.D.

Cornell University 2014

Arbuscular mycorrhizal fungi (AMF) are obligate biotrophs that form symbiotic associations with the roots of terrestrial plants. All major lineages of AMF harbour cytoplasmic endobacteria related to the Mollicutes class, referred to as the Mollicutes/mycoplasma-related endobacteria (MRE). There is virtually nothing known about the biology of MRE, nor the nature of their association of with AMF. The research outlined in this thesis is the first biological study of MRE. The first study is the official classification of MRE, based on our phylogenetic analysis and *in vitro* cultivation attempts. Our study revealed that MRE consist of many types, with MRE from each host species displaying varying degrees of genetic diversity. We used the genetically uniform MRE population of the AMF *Rhizophagus clarus*, as the first representative for official classification of MRE; *Candidatus* Glomeriplasma moenium gen. nov., sp. nov. was proposed for this MRE population. The second study is a metagenomic study on three MRE populations. We discovered that MRE genomes are highly plastic, with evidence of numerous chromosomal rearrangements. Their minimal genome indicates complete metabolic dependence on their hosts. The MRE genome also contained many genes acquired from the AMF through horizontal gene transfer, including SUMO proteases, likely used to alter SUMOylation levels in their hosts, which has been shown to increase endobacterial fitness in other systems. The extent of MRE genome erosion, along with the large number of horizontally acquired fungal genes, suggests a

high degree of coevolution between the partners, and highlights the significance of MRE in AMF biology. The third study is an investigation of the molecular evolution of MRE. Muller's ratchet is an evolutionary theory that predicts that asexual finite populations (e.g. endobacterial populations) will eventually lead to population extinction due to the continuous fixation of deleterious mutations combined with the random loss of the most fit individuals. We found that MRE escapes Muller's ratchet through the maintenance of recombination machinery that creates genetic diversity in their population. MRE are unique in this light, as endobacteria are generally unable to prevent the erosion of their recombination machinery.

BIOGRAPHICAL SKETCH

Mizue Naito was born on February 4th, 1983 in Kyoto, Japan. She attended Montessori school from the age of 3, until her family moved to Toronto, Canada in 1988, where she continued her education. She received her H.B.Sc. from the University of Toronto in 2006, where she specialized in Medical Genetics & Microbiology. For her undergraduate research thesis, she conducted research in the Mogridge Lab, on the biochemical properties of the anthrax toxin receptor, ANTXR1. She then received her M.Sc. from the University of British Columbia in 2008, from the Department of Microbiology & Immunology. For her Master's thesis, she worked in the Gaynor Lab, and conducted research on the roles of surface sugar moieties on the pathogenesis of *Campylobacter jejuni*, an enteric human pathogen. She returned to Toronto after her graduation and worked as a Research Technologist at the Ciruna Lab at the Hospital for Sick Children for a year and a half, conducting research on novel molecular methods for site-directed mutagenesis in the zebrafish model system.

Mizue realized she wanted to further her career, and so she applied to and enrolled in the Microbiology program at Cornell University in 2010. She joined the Pawlowska Lab for her doctorate research, where she studied the biology and evolution of the novel endobacteria of mycorrhizal fungi. After her graduation, she plans on moving back to Toronto, to continue her career as a Microbiologist.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my major advisor, Teresa Pawlowska. She has been a wonderful mentor, and always provided both guidance and freedom in my projects. She has also shown great support outside of the lab, and I will always be grateful for her kindness and admire her genuine care for her students. I would also like to thank my minor advisors, Michael Stanhope and Gillian Turgeon, for their guidance throughout my Ph.D. career.

I would also like to thank my dearest friends that I have met in Cornell, especially Basak Belet and Jeanne Quirit. Both Cornell and Ithaca gave us quite the culture shock, and I would not have been able to get past all of the negativity that surrounds this place without their support and friendship. My family and friends back home, especially my best friend Aprile Manjoo, also deserve my many thanks for always supporting me and listening to my never-ending rants about life in the U.S.

Finally, I would like to thank Matthew Bucemi, for always being kind, thoughtful, and caring towards me. I am extremely grateful to have you in my life, and am looking forward to our next chapter together.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC_c: akaike information criterion, corrected for finite sample size
AMF: arbuscular mycorrhizal fungi
ANK: ankyrin repeats
AT: adenine-thymine
BEAST: Bayesian evolutionary analysis sampling trees
BHI: brain heart infusion
BLASTp: basic local alignment search tool, protein sequence
BQH: black queen hypothesis
CDS: coding deoxyribonucleic acid sequence
CRISPR: clustered regularly interspaced short palindromic repeats
bp: base pairs
DNA: deoxyribonucleic acid
FISH: fluorescent *in situ* hybridization
GARD: genetic algorithm for recombination detection
GC: guanine-cytosine
GIMAP: guanosine-5'-triphosphatase of immunity-associated proteins
GTP: guanosine-5'-triphosphate
GTR: general time reversal
HET: heterokaryon incompatibility
HGT: horizontal gene transfer
IAN: immune-associated nucleotide-binding protein
IMG: integrated microbial genomes
INVAM: international culture collection of vesicular arbuscular mycorrhizal fungi
KAAS: Kyoto encyclopaedia of genes and genomes automatic annotation server
kb: kilobases
KEGG: Kyoto encyclopaedia of genes and genomes
KH: Kishino-Hasegawa
KO: Kyoto encyclopaedia of genes and genomes orthology
LCA: last common ancestor
LRR: leucine rich repeats
M: minimal/modified White's
Mb: megabases
MRE: Mollicutes/Mycoplasma-related endobacteria
MRE-CE: Mollicutes/Mycoplasma-related endobacteria from *Claroideoglomus etunicatum*
MRE-RC: Mollicutes/Mycoplasma-related endobacteria from *Rhizophagus clarus*
MRE-RV: Mollicutes/Mycoplasma-related endobacteria from *Racocetra verrucosa*
MSR: modified Strullu-Romand
MUSCLE: multiple sequence comparison by log-expectation
ORF: open reading frame
PCR: polymerase chain reaction
PPLO: pleuropneumonia-like organisms
RNA: ribonucleic acid
rRNA: ribosomal ribonucleic acid
SSR: simple sequence repeats

SUMO: small ubiquitin-related modifier
SUMOfi: small ubiquitin-related modifier motif finder
TCA: tricarboxylic acid cycle
tRNA: transfer ribonucleic acid
WAG: Whelan and Goldman

CHAPTER 1

INTRODUCTION

1.1.0 - ARBUSCULAR MYCORRHIZAL FUNGI (AMF)

Arbuscular mycorrhizal fungi (AMF) comprise the monophyletic phylum, Glomeromycota, and consist of over 200 species (Redecker et al 2013, Schussler and Walker 2010). AMF are obligate biotrophs, and form symbiotic association with the roots of up to 80% of terrestrial plant species; this association is believed to be the most common multi-partner symbiosis on the planet (Wang and Qiu 2006). Symbiotically associated AMF connect their host plants' root system to their extensive hyphal network in the rhizosphere, which can be in excess of 100 metres of hyphae per cm^3 of soil (Miller et al 1995), and provide an increased uptake of nutrients to the plants (Parniske 2008). In return for the increased nutrient and water uptake, the plants supply AMF with up to 20% of their photosynthetically derived carbon energy (Bago et al 2000, Fellbaum et al 2012). Thus, AMF are a key player in terrestrial ecosystems and contribute significantly to the global nutrient cycle. In addition to providing increased nutrient uptake, AMF have also been shown to promote floristic diversity (Grime et al 1987), and improve plant fitness against pathogens (Liu et al 2007).

Fossil records indicate that the AMF-plant symbiosis is over 400 million years old (Remy et al 1994). It is widely believed that this AMF-plant symbiosis facilitated the transition of plants from the aquatic to the terrestrial environment. Owing perhaps to their long symbiotic history, and their lack of host plant specificity, AMF are currently found in almost all terrestrial ecosystems, from the tropics through temperate forests, deserts, and grasslands to agroecosystems (Sturmer 2012). AMF are now emerging as the key player in the future of

sustainable agriculture, with the goal of optimizing AMF-crop symbiosis to replace the use of non-renewable resources (Gianinazzi et al 2010).

1.1.1 - Life cycle of AMF

AMF are an unusual group of fungi; their hyphae are asceptate, and contain hundreds of haploid nuclei within the same cytoplasm. They are considered to be asexuals, with no evidence of sexual reproduction, despite the presence of mating gene homologues in both the genome and transcripts of *Rhizophagus irregularis*, the only sequenced AMF to date, and in low-coverage genome surveys of closely related species (Halary et al 2011, Halary et al 2013, Tisserant et al 2012, Tisserant et al 2013). Despite its asexual lifestyle, evidence of infrequent genetic exchanges exists in AMF, believed to be the consequence hyphal anastomosis between different AMF individuals (Chagnon 2014, Croll et al 2009, Croll and Sanders 2009, den Bakker et al 2010).

For AMF to complete their life cycle, they depend on a plant host to provide them with much needed photosynthate energy. AMF spores in the rhizosphere are capable of germination and hyphal extension (limited by the available lipid resources of the spore), known as the presymbiotic phase. They are able to grow towards potential hosts by perception of plant root exudates, such as strigolactones (Besserer et al 2006). The AMF also release fungal factors, including the Myc factor, that, in turn, induce transcriptional activation of symbiosis-related genes in the nearby host (Kloppholz et al 2011, Nadal and Paszkowski 2013, Olah et al 2005). When AMF physically encounter a host root, their hyphae swell to form a hyphopodium, which allows for hyphal penetration into the root epidermal cell (Genre et al 2005). The hyphae will then continue to grow until they reach the inner root cortex, where they form arbuscules (their

namesake, from the Latin, *arbuscula*, meaning small tree), which are extensive branching of intracellular hypha (Bonfante and Genre 2010). The arbuscules are believed to be the main site of nutrient exchange between the AMF and their hosts. With the successful association and symbiotic exchange with a plant host, new AMF spores are formed outside of the plant root, and the life cycle is complete.

1.1.2 - The endobacteria of AMF

For almost half a century, AMF have been known to harbour vertically transmitted endobacteria in their cytoplasm, based on observations with electron microscopy (Macdonald and Chandler 1981, Mosse 1970, Protsenko 1975). It is now known that there are two distinct types of endobacteria: *Candidatus* Glomeribacter gigasporarum, a *Burkholderia*-related species, and a yet unnamed bacteria, referred to as Mollicutes/mycoplasma-related endobacteria (MRE) (Bianciotto et al 2003, Naumann et al 2010).

Ca. Glomeribacter gigasporarum are nonessential endosymbionts of AMF, despite the fact that the association is over 400 million years old (Mondo et al 2012). The rod-shaped bacteria are compartmentalized in cytoplasmic vacuoles of AMF, and their host range is restricted to the Gigasporaceae family (Bianciotto et al 2003). Based on a study in which *Gigaspora margarita*, an AMF species that normally harbour *Ca.* Glomeribacter gigasporarum, was successfully cured of its endobacteria, it became evident that this bacterial-fungal association is mutualistic; the AMF host provides both niche and nutrients for the bacteria, and the bacteria provide increased ability of presymbiotic growth for the fungi (Lumini et al 2007). Though the precise mechanisms of these interactions remain unclear, the 1.7 Mb genome of the

bacterium suggests that a mosaic of gene homologues involved in symbiotic, pathogenic, and free-living species are present (Ghignone et al 2012).

The second class of endobacteria, the MRE, is the focus of this study. Electron micrographs have revealed that these bacteria are coccoid in shape, around 0.5 μm in diameter, and unlike *Ca.* Glomeribacter gigasporarum, live freely in the AMF cytoplasm without any compartmentalization (Naumann et al 2010). MRE are also found associated with almost every lineage of the AMF host, and present in many AMF worldwide (Toomer et al 2014). A survey of 16S rRNA sequences of MRE revealed that they are a member of the Mollicutes (Naumann et al 2010). Virtually no other information is known about the MRE, such as their role in AMF (parasite or mutualist?), their exact classification, or their biological capabilities.

1.1.3 - AMF species used in this study

In order to study the biology of MRE, three AMF host species were selected: *Claroideoglomus etunicatum* OT135, *Racocetra verrucosa* VA103A, and *Rhizophagus clarus* NB112A. *C. etunicatum* is a member of the family Claroideoglomeraceae, and *R. clarus* is a member of the family Glomeraceae. Both are members of the order Glomerales. *R. verrucosa* is a member of the Gigasporaceae family, in the order Diversisporales (Redecker et al 2013, Schussler and Walker 2010).

The AMF species, *Rhizophagus irregularis*, a member of the Glomeraceae family and Glomerales order, is also mentioned or used in this study. This AMF species does not harbour any endobacteria. In addition, *R. irregularis* is the only sequenced AMF species to date, and thus can provide valuable information on the genetic potential of AMF (Tisserant et al 2013).

1.2.0 - BACTERIAL ENDOSYMBIONTS

Endosymbiosis is a phenomenon where one organism lives within another organism, without specifying the net fitness outcome to the host. Endosymbiosis, and particularly bacterial endosymbiosis, where the intracellular partner is of bacterial origin, has had profound impacts on the evolution of planetary species; mitochondria and chloroplasts are the products of ancient bacterial endosymbionts that allowed for hosts to survive in the presence of oxygen and convert light energy into chemical energy (Dyall et al 2004). As bacterial endosymbionts and their hosts are often members of distinct domains of life, the association allows for novel biological capabilities, and survival in environments that would otherwise be impossible for either species alone (Wernegreen 2012).

The biological and biochemical benefit of bacterial endosymbionts are vast and specialized for each host. For example, a marine gutless oligochaete worm, *Olavius algarvensis*, harbours a consortium of sulphate-reducing and sulphide-oxidizing bacteria that produce energy for the host, allowing for the devolution of worm's entire mouth, gut, and nephridial excretory organs (Dubilier et al 2001, Kleiner et al 2012). Insects are particularly known for harbouring various bacterial endosymbionts, most commonly due to their diet of plant sap that lacks essential nutrients, which the symbionts can provide. An example of this is *Candidatus* Zinderia insecticola and *Candidatus* Sulcia muelleri, both endosymbionts of the spittlebug *Clastoptera arizonana*. *Ca.* Zinderia insecticola are capable of synthesizing histidine, methionine and tryptophan, while *Ca.* Sulcia muelleri synthesize 7 other essential amino acids that complement those not produced by *Ca.* Zinderia insecticola, and together they provide the amino acids necessary for both the symbionts themselves and for their insect host (McCutcheon and Moran 2010). Many legume plants are also known to associate with various nitrogen-fixing bacteria,

collectively known as rhizobia, to form root nodules. The bacteria reside in membrane-bound vesicles in the nodules, where they reduce atmospheric nitrogen for the plants, while receiving carbon energy in return (Kondorosi et al 2013). This important interaction is highly species specific (both plant and rhizobia species), and the genes involved in this symbiotic association have been extensively studied (Tian et al 2012). Bacterial endosymbionts have also been known to confer protection for their hosts against other predators. The aphid, *Myzus persicae*, is one such example, where the presence of the endosymbiont *Regiella insecticola* provided resistance against the parasitoid, *Aphidius colemani*, and this protection was further shown to be transferable to other aphid species (Vorburger et al 2010). The most unusual case of bacterial endosymbiosis is likely to be that of the citrus mealybug, *Planococcus citri*, which hosts the bacterial endosymbiont, *Candidatus* Tremblaya princeps that, in turn, hosts its own bacterial endosymbiont, *Candidatus* Moranella endobia, the only known bacterial endosymbiont of another bacteria (McCutcheon and von Dohlen 2011, von Dohlen et al 2001).

1.2.1 - Bacterial endosymbionts of fungi

Despite numerous studies on endosymbionts, there are relatively few studies on bacterial endosymbionts of fungi. The only well studied and characterized association is that of *Rhizopus microsporus,* a plant pathogenic fungus best known for causing rice seedling blight, and its endosymbiont, *Burkholderia rhizoxinica*. In this association, the bacteria produce the toxin rhizoxin for their fungal host, which allows the fungus to access nutrients from the necrotized plants (Partida-Martinez and Hertweck 2005). Furthermore, the reproduction of the fungal host was shown to be completely dependent on its endobacteria, where the absence of bacteria prevented asexual reproduction of the host (Partida-Martinez et al 2007). The genome of *B.*

rhizoxinica has been sequenced, and at 3.75 Mb, encodes for genes specialized in the uptake of fungal metabolites, along with the type III secretion system and effectors that are believed to be important for the bacterial-fungal interaction (Lackner et al 2011a, Lackner et al 2011b, Lackner et al 2011c).

Other interactions between bacteria and fungi are just beginning to emerge. The relationship of Gigasporaceae AMF and *Ca.* Glomeribacter gigasporarum is described in Chapter 1.1.2. Other interactions include two Mucoromycotina species, *Mortierella elongata* and *Mortierella alpina,* both shown to harbour endobacteria (Kai et al 2012, Sato et al 2010). The nature of these relationships remains to be discovered. It is clear that bacterial-fungal interactions are abundant in nature, and likely have important effects on both the biology and ecology of the organisms.

1.3.0 - BACTERIA OF THE CLASS MOLLICUTES

Mollicutes are a class of bacteria, under the phylum Tenericutes. Bacteria of this class are distinguished by the lack of a cell wall, despite the fact that they are closely related to other Gram-positive bacteria. For this reason, the name Mollicutes (Latin *mollis* = soft, pliable; *cutis* = skin) was given to this class (Edward and Freundt 1967). It is now known that their lack of cell wall is due to reductive evolution, along with other key features of this class, such as small genome size, small number of rRNA operons and tRNA genes, fastidious growth conditions, and limited metabolic capabilities (Bove 1993, Pollack 2002). Furthermore, the genomes of Mollicutes are AT-rich, and this pressure has caused the switch of the UGA opal stop codon to encode for tryptophan (Sirand-Pugnet et al 2007). This switch is accompanied by the loss of the

peptide release factor 2, and the gain of $tRNA^{trp}$ that harbour a U*CA anticodon (Citti et al 1992, Inagaki et al 1993).

The genomes of Mollicutes are also known to be highly plastic. Attempts to align whole genomes of Mollicutes species have shown a general lack of gene synteny in this group, and long regions with conserved gene synteny are only seen between strains of some species (Sirand-Pugnet et al 2007). A phylogenetic analysis of conserved gene clusters of microbial genomes has also revealed that the Mollicutes have one of the lowest numbers of predicted clusters (Zheng et al 2005). Chromosomal rearrangements are frequently observed in this class, due to a variety of mechanisms including mobile elements, large chromosomal duplications, the presence of highly repetitious sequences that serve as sites of homologous recombination, and even the presence of bacteriophages (Ku et al 2013, Mrazek 2006, Rocha and Blanchard 2002, Sugio and Hogenhout 2012). It is now increasingly clear, that despite their small genome and the previous belief of their limited capabilities, the Mollicutes have evolved numerous mechanisms to maintain a highly plastic genome, and with it the ability to generate novel functions and survival in various niches.

1.3.1 - Bacteria of the order Entomoplasmatales -- Mesoplasma and Spiroplasmas

The Entomoplasmatales is a small order of the Mollicutes, and includes the bacterial species, *Entomoplasma*, *Mesoplasma*, and *Spiroplasma*. The *Mesoplasma* species are found associated with both insects and plants. They differ from other Mollicutes by their lack of need for sterol as a requirement for growth (Tully et al 1994). Furthermore, they generally lack pathogenic potential, and are non-motile (Baby et al 2013). Interestingly, *Mesoplasma florum*, the most studied species of this genus, possesses a unique form of riboswitch variant that binds

2'-deoxyguanosine (Kim et al 2007). This variant is believed to be a product of mutation accumulation, creating a unique ligand specificity found only in this particular bacterial species.

The *Spiroplasma* species are a group of motile, helical bacteria, and are associated with insects and plants (Gasparich 2010). Several species are capable of growth in both plant phloem and insect hosts, and plant surfaces are important in the horizontal transmission of the species among insect hosts (Clark et al 1987). Though most *Spiroplasmas* are non-pathogenic in nature, certain species have been shown to cause disease in plants, insects, or arthropod hosts (Bai and Hogenhout 2002, Clark et al 1985, Nunan et al 2005). Some maternally transmitted *Spiroplasma* species can also cause sex-ratio distortion in *Drosophila* flies, where male species are selectively killed during embryogenesis to promote bacterial transmission (Harumoto et al 2014, Williamson et al 1999). The genomes of *Spiroplasma citri, Spiroplasma kunkelii,* and *Spiroplasma melliferum* contain numerous mobile elements and full-length prophages of plectrovirus integrated into the chromosomes (Carle et al 2010, Melcher et al 1999). The plectroviral fragments have been linked to chromosomal rearrangements in these bacteria, and are one of the causative agents of their extreme genome instability (Ku et al 2013, Ye et al 1996).

1.3.2 - Bacteria of the order Mycoplasmatales -- Ureaplasma and Mycoplasma

The Mycoplasmatales is an order of the Mollicutes, and includes the bacterial species, *Candidatus* Hepatoplasma, *Ureaplasma*, and the most well studied Mollicutes genus, the *Mycoplasma. Ureaplasma* species is a unique member of the Mollicutes that can hydrolyze urea for energy production (Glass et al 2000). Though generally a harmless commensal, it is also implicated in numerous complications in human pregnancies and complications and mortalities in neonates (Viscardi 2010, Volgmann et al 2005). Similar to some other human pathogens of

the Mollicutes class, *Ureaplasma* species express surface antigens that are highly variable, that enable their evasion from host immune systems (Waites et al 2005, Zheng et al 1992). Furthermore, extensive horizontal gene transfer is evident in the genomes of *Ureaplasma* species, indicating that the species is rapidly evolving (Paralanov et al 2012).

The *Mycoplasma* species is the most well studied genus of the Mollicutes class of bacteria. The genus includes more than 100 species, many of which are pathogenic for a range of eukaryotic hosts, including humans and animals of agricultural and veterinary importance (Citti and Blanchard 2013). *Mycoplasma pneumoniae*, for instance, is a common cause of upper and lower respiratory tract infections in humans, and may also play a role in chronic human infections such as rheumatoid arthritis and asthma (Waites et al 2008). Another species of human clinical importance is *Mycoplasma genitalium*, which causes urethritis in men and a variety of reproductive tract syndromes in women, including infertility (McGowin and Anderson-Smits 2011). Non-human pathogens include *Mycoplasma alligatoris*, the causative agent of lethal necrotizing pneumonia in adult alligators (Brown et al 2004), and *Mycoplasma leachii*, the causative agent of polyarthritis in calves and mastitis in cows (Chang et al 2011).

Mycoplasmas have also gained scientific interest due to their extremely small genomes. At 580 kb, *M. genitalium* has the smallest genome of a free-living bacterium that is not an obligate endosymbiont (Fraser et al 1995). For this reason, *M. genitalium* is frequently used as a basis for study of the "minimal cell", or the minimal set of genes essential for life (Glass et al 2006, Hutchison et al 1999). Furthermore, *M. genitalium* was used as the blueprint for the first fully synthesized genome (Gibson et al 2008), followed by the synthesis of a 1.08 Mb genome of *Mycoplasma mycoides*, transplanted into a *Mycoplasma capricolum* recipient cell, to create the first cell controlled only by a synthetic chromosome (Gibson et al 2010).

Despite their small genomes, it is clear that the mycoplasmas are well adapted, and not restricted in their biological capabilities. This seemingly contradictory phenomenon is likely due to their highly plastic genome structure, and their extremely rapid evolutionary rates. For example, *Mycoplasma gallisepticum* is known to have one of the fastest evolving genomes in bacteria, at 0.8-1.2x10^5 nucleotide substitutions per site per year (Delaney et al 2012). This capacity is believed to be the reason behind the bacteria's ability for host switching and adaptation; *M. gallisepticum* is believed to have jumped hosts from poultry to house finches in the early 90's, leading to the death of 225 million finches within 3 years (Citti and Blanchard 2013, Ley et al 1996). Horizontal gene transfer is also widely seen in many *Mycoplasma* species, including *M. leachii*, whose genome suggest a mosaic of sequences derived from *M. mycoides* subsp. *mycoides*, and *M. mycoides* subsp. *capri*, both ruminant pathogens (Tardy et al 2009). Furthermore, plasmids in the mycoides group also display mosaic structures, indicative of recombination events and frequent plasmid transmission between species sharing a common host (Breton et al 2012). Numerous studies on *Mycoplasma* cell surface structures have also revealed that many species are capable of frequent changes in expression and alteration of their antigenic makeup of surface proteins, with methods including intragenomic recombination and slipped-strand mispairing events (Ma et al 2007, Mrazek 2006, Shaw et al 2012). Mobile elements are also frequent in some species of *Mycoplasmas*, and have been shown to be the cause of their genome plasticity (Dordet Frisoni et al 2013, Qi et al 2012).

1.3.3 - Mollicutes species capable of intracellular survival

Some species of Mollicutes are capable of invasion into host cells and intracellular survival. *Mycoplasma penetrans*, *M. pneumoniae*, and *M. genitalium* have been shown to

replicate intracellularly in continuous passages of human cell lines (Dallo and Baseman 2000). A unique form of intracellular survival has been adapted by *Mycoplasma hominis*, found in human genital tracts. *M. hominis* forms a symbiotic relationship with *Trichomonas vaginalis*, a parasitic protozoan responsible for trichomoniasis, by entering the protozoan cell, which provides a protective niche against the human immune system (Dessi et al 2005, Vancini and Benchimol 2008). *T. vaginalis* seems to gain an increase in cytosolic ornithine and putrescine concentration from *M. hominis*, potentially beneficial for downstream energy production (Morada et al 2010). Though the mechanisms of host cell invasions are still mainly unclear, it is evident that intracellular survival has been adapted independently many times in various Mollicutes species, and some have evolved to become a mutualistic interaction with their hosts.

1.4.0 - GENOME EVOLUTION OF ENDOSYMBIONTS

Bacterial endosymbionts that are strictly vertically transmitted undergo a unique genomic evolution that is not seen in those of free-living status. The most prominent feature is extreme genome reduction, where bacterial genomes are typically reduced to less than 1 Mb (Moran et al 2008). For instance, the smallest genome of an endosymbiont sequenced to date is *Candidatus* Nasuia deltocephalinicola, the endosymbiont of the insect pest, *Macrosteles quadrilineatus,* with a genome of only 112 kb (Bennett and Moran 2013). The genes lost in reduced endobacterial genomes come from all functional categories, including those involved in DNA repair, while genes that are maintained are those involved in essential cellular processes such as DNA replication, transcription, and protein translation (Moran et al 2008). Other common genomic features include rapid sequence evolution, where endobacteria always appear on long branches

of phylogenetic trees, codon reassignments, and biases in nucleotide composition where adenine and thymine are favoured (McCutcheon and Moran 2012).

All of the above features are due to the endosymbionts' distinct lifestyle of vertical transmission through host generations. Due to a combination of factors such as asexuality of bacteria, and bottlenecks at each transmission event, the endobacterial populations have small effective sizes and experience weak purifying selection, resulting in the accumulation of slightly deleterious mutations (Moran 1996). Accumulation of mutations can lead to inactivation and deletion of genes, resulting in a reduced and degenerative genome (Nilsson et al 2005). Furthermore, as the degenerative population is restricted to specific hosts, with no influx of novel genotypes, the population will continue to lose its best fit individuals; this phenomenon is known as Muller's ratchet (after the evolutionist, Hermann Joseph Muller). The Muller's ratchet model predicts the eventual extinction of small asexual populations (Crow 2005). This population structure is also the cause of accelerated sequence evolution in endosymbionts, further strengthened by the loss of DNA repair genes seen in such populations. The accelerated evolution and the fixation of deleterious mutations leads to the expression of many proteins with lowered predicted thermal stability (McCutcheon and Moran 2012, van Ham et al 2003). For this reason, endobacteria also constitutively express elevated levels of heat shock proteins and chaperonins (Fares et al 2002, Tokuriki and Tawfik 2009). Such examples of extreme genome erosion and reduction only occur in beneficial endosymbionts, since co-adaptation with the host is a crucial component in enabling the massive loss of genomic components (McCutcheon and Moran 2012).

It has been assumed for many years that genome reduction of endosymbionts follows a strict evolutionary path: early stages are characterized by an overabundance of mobile elements,

pseudogenes, and multiple genomic rearrangements, while latter stages, or ancient endosymbionts are characterized by the lack of mobile elements, loss of most pseudogenes, and a relatively stable yet continuously reducing genome (McCutcheon and Moran 2012, Ochman and Davalos 2006). However, recent studies and the increased sequencing of endobacterial genomes have revealed that this may not be the case. Genomes of *Wolbachia* species, reproductive parasites of insects, have revealed that despite their long-term association with their insect host and reduced genomes, the species still harbour an over-abundance of mobile elements, and extensive genomic insertions (Brelsfoard et al 2014, Leclercq et al 2011). Analysis of the louse endosymbiont, *Candidatus* Riesia species, revealed a reduction in mutation accumulation, leading to a slowdown of the effects of Muller's ratchet (Allen et al 2009). Extensive genome rearrangement has also been documented in the genomes of *Portiera* species, the endosymbiont of whiteflies, likely due to recombination events across identical repeats found in the bacterial genome (Sloan and Moran 2013). A novel evolutionary model has also been suggested, where reductive evolution is a byproduct of natural selection, rather than genetic drift (Morris et al 2012). This model is termed the Black Queen Hypothesis, and may also explain some of the reduced genomes seen in endosymbionts. It is clear that the genome evolution in endosymbionts is influenced by multiple factors, and thus the trajectory of the endobacterial population may not be as simple and predictable as once believed. The evolution of endosymbiotic populations not only has profound effects on the population itself, but also for the host species that harbour them.

1.5.0 - CONTENT OF DISSERTATION

This dissertation investigates a novel endobacteria of AMF, known as MRE. The study is separated into 3 chapters.

Chapter 2 is the classification and taxonomic work on the MRE. These bacteria form a novel clade of the Mollicutes class. This investigation has led to the proposal of a novel genus and species, *Candidatus* Glomeriplasma moenium, under a novel family, *Glomeriplasmataceae*.

Chapter 3 is a metagenomic study of three MRE populations extracted from three distinct AMF host species. This study is the first study looking into the biology and function of the MRE. This investigation has revealed that the MRE genomes are highly plastic, undergoing extensive chromosomal rearrangements. Furthermore, all three populations have a high number of genes that encode for proteins that interact with host AMF proteins. Many genes in the MRE genome have fungal origins, pointing to extensive horizontal gene transfer events from the AMF host to the MRE; some of these genes are involved in affecting the host cytoplasmic environment, resulting in an outcome that is likely favourable for MRE stability in the host cytoplasmic niche. The metagenomic study has revealed that despite evidence of genomic erosion, MRE genotypes with high levels of fungal genes, or genes that interact with eukaryotic proteins are being selected/maintained. We believe this suggests that the MRE-AMF association is genetically tightly knit, and is likely an obligate association for the AMF species that harbour them.

Chapter 4 is an analytical study on the genome structure of the MRE, and its molecular evolution. Vertically transmitted endobacterial populations have unique evolutionary forces that shape their genomic structures, unseen in free-living bacteria. Such forces lead to the continuous degeneration of the population's genome, and the random loss of fit individuals from the

population, an evolutionary model known as Muller's ratchet. We found that MRE populations retain an unusual genome structure, with the retention of genes that maintain diversity in the population. The MRE population is unique in their ability to maintain "sexuality" to prevent the erosion of their population through Muller's ratchet. This rare adaptation by MRE, and its implication in endobacterial evolution is discussed.

This dissertation outlines the first extensive study on the biology of the MRE. MRE are members of the Mollicutes class, whose other representatives all have important interactions with eukaryotic hosts. MRE reside only in the cytoplasm of AMF species, distributed worldwide. As AMF are found in their own symbiotic association with the majority of land plants, and are key players in the terrestrial ecosystem and global nutrient cycling, the role of the MRE as a hidden player in this important environmental community is an important research area. AMF are also important members in the future of sustainable agriculture, and thus the study of MRE is required for the optimal interaction of MRE-AMF-agricultural plant symbiosis.

REFERENCES

Allen JM, Light JE, Perotti MA, Braig HR, Reed DL (2009). Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet. *PLoS One* **4:** e4969.

Baby V, Matteau D, Knight TF, Rodrigue S (2013). Complete Genome Sequence of the Mesoplasma florum W37 Strain. *Genome Announc* **1**.

Bago B, Pfeffer PE, Shachar-Hill Y (2000). Carbon metabolism and transport in arbuscular mycorrhizas. *Plant Physiol* **124:** 949-958.

Bai X, Hogenhout SA (2002). A genome sequence survey of the mollicute corn stunt spiroplasma Spiroplasma kunkelii. *FEMS Microbiol Lett* **210:** 7-17.

Bennett GM, Moran NA (2013). Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biol Evol* **5:** 1675-1688.

Besserer A, Puech-Pages V, Kiefer P, Gomez-Roldan V, Jauneau A, Roy S *et al* (2006). Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. *PLoS Biol* **4:** e226.

Bianciotto V, Lumini E, Bonfante P, Vandamme P (2003). 'Candidatus glomeribacter gigasporarum' gen. nov., sp. nov., an endosymbiont of arbuscular mycorrhizal fungi. *Int J Syst Evol Microbiol* **53:** 121-124.

Bonfante P, Genre A (2010). Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat Commun* **1:** 48.

Bove JM (1993). Molecular features of mollicutes. *Clin Infect Dis* **17 Suppl 1:** S10-31.

Brelsfoard C, Tsiamis G, Falchetto M, Gomulski LM, Telleria E, Alam U *et al* (2014). Presence of extensive Wolbachia symbiont insertions discovered in the genome of its host Glossina morsitans morsitans. *PLoS Negl Trop Dis* **8:** e2728.

Breton M, Tardy F, Dordet-Frisoni E, Sagne E, Mick V, Renaudin J *et al* (2012). Distribution and diversity of mycoplasma plasmids: lessons from cryptic genetic elements. *BMC Microbiol* **12:** 257.

Brown DR, Zacher LA, Farmerie WG (2004). Spreading factors of Mycoplasma alligatoris, a flesh-eating mycoplasma. *J Bacteriol* **186:** 3922-3927.

Carle P, Saillard C, Carrere N, Carrere S, Duret S, Eveillard S *et al* (2010). Partial chromosome sequence of Spiroplasma citri reveals extensive viral invasion and important gene decay. *Appl Environ Microbiol* **76:** 3420-3426.

Chagnon PL (2014). Ecological and evolutionary implications of hyphal anastomosis in arbuscular mycorrhizal fungi. *FEMS Microbiol Ecol*.

Chang JT, Liu HJ, Yu L (2011). Mycoplasma leachii sp. nov. in calves, China. *Emerg Infect Dis* **17:** 1772-1773.

Citti C, Marechal-Drouard L, Saillard C, Weil JH, Bove JM (1992). Spiroplasma citri UGG and UGA tryptophan codons: sequence of the two tryptophanyl-tRNAs and organization of the corresponding genes. *J Bacteriol* **174:** 6471-6478.

Citti C, Blanchard A (2013). Mycoplasmas and their host: emerging and re-emerging minimal pathogens. *Trends Microbiol* **21:** 196-203.

Clark TB, Whitcomb RF, Tully JG, Mouches C, Saillard C, Bove JM *et al* (1985). *Spiroplasma melliferum*, a new species from the honeybee (*Apis mellifera*). *Int J Syst Bacteriol* **35:** 296-308.

Clark TB, Henegar RB, Rosen L, Hackett KJ, Whitcomb RF, Lowry JE *et al* (1987). New spiroplasmas from insects and flowers: isolation, ecology, and host association. *Isr J Med Sci* **23:** 687-690.

Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ *et al* (2009). Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus Glomus intraradices. *New Phytol* **181:** 924-937.

Croll D, Sanders IR (2009). Recombination in Glomus intraradices, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evol Biol* **9:** 13.

Crow JF (2005). Timeline: Hermann Joseph Muller, evolutionist. *Nat Rev Genet* **6:** 941-945.

Dallo SF, Baseman JB (2000). Intracellular DNA replication and long-term survival of pathogenic mycoplasmas. *Microb Pathog* **29:** 301-309.

Delaney NF, Balenger S, Bonneaud C, Marx CJ, Hill GE, Ferguson-Noel N *et al* (2012). Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, Mycoplasma gallisepticum. *PLoS Genet* **8:** e1002511.

den Bakker HC, Vankuren NW, Morton JB, Pawlowska TE (2010). Clonality and recombination in the life history of an asexual arbuscular mycorrhizal fungus. *Mol Biol Evol* **27:** 2474-2486.

Dessi D, Delogu G, Emonte E, Catania MR, Fiori PL, Rappelli P (2005). Long-term survival and intracellular replication of Mycoplasma hominis in Trichomonas vaginalis cells: potential role of the protozoon in transmitting bacterial infection. *Infect Immun* **73:** 1180-1186.

Dordet Frisoni E, Marenda MS, Sagne E, Nouvel LX, Guerillot R, Glaser P *et al* (2013). ICEA of Mycoplasma agalactiae: a new family of self-transmissible integrative elements that confers conjugative properties to the recipient strain. *Mol Microbiol* **89:** 1226-1239.

Dubilier N, Mulders C, Ferdelman T, de Beer D, Pernthaler A, Klein M *et al* (2001). Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* **411:** 298-302.

Dyall SD, Brown MT, Johnson PJ (2004). Ancient invasions: from endosymbionts to organelles. *Science* **304:** 253-257.

Edward DG, Freundt EA (1967). Proposal for *Mollicutes* as name of the class established for the order *Mycoplasmatales*. *Int J Syst Bacteriol* **17:** 267-268.

Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E (2002). Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature* **417:** 398.

Fellbaum CR, Gachomo EW, Beesetty Y, Choudhari S, Strahan GD, Pfeffer PE *et al* (2012). Carbon availability triggers fungal nitrogen uptake and transport in arbuscular mycorrhizal symbiosis. *Proc Natl Acad Sci U S A* **109:** 2666-2671.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD *et al* (1995). The minimal gene complement of Mycoplasma genitalium. *Science* **270:** 397-403.

Gasparich GE (2010). Spiroplasmas and phytoplasmas: microbes associated with plant hosts. *Biologicals* **38:** 193-203.

Genre A, Chabaud M, Timmers T, Bonfante P, Barker DG (2005). Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in Medicago truncatula root epidermal cells before infection. *Plant Cell* **17:** 3489-3499.

Ghignone S, Salvioli A, Anca I, Lumini E, Ortu G, Petiti L *et al* (2012). The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. *ISME J* **6:** 136-145.

Gianinazzi S, Gollotte A, Binet MN, van Tuinen D, Redecker D, Wipf D (2010). Agroecology: the key role of arbuscular mycorrhizas in ecosystem services. *Mycorrhiza* **20:** 519-530.

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J *et al* (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319:** 1215-1220.

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA *et al* (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329:** 52-56.

Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH (2000). The complete sequence of the mucosal pathogen Ureaplasma urealyticum. *Nature* **407:** 757-762.

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M *et al* (2006). Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* **103:** 425-430.

Grime JP, Mackey JML, Hillier SH, Read DJ (1987). Floristic Diversity in a Model System Using Experimental Microcosms. *Nature* **328:** 420-422.

Halary S, Malik SB, Lildhar L, Slamovits CH, Hijri M, Corradi N (2011). Conserved meiotic machinery in Glomus spp., a putatively ancient asexual fungal lineage. *Genome Biol Evol* **3:** 950-958.

Halary S, Daubois L, Terrat Y, Ellenberger S, Wostemeyer J, Hijri M (2013). Mating type gene homologues and putative sex pheromone-sensing pathway in arbuscular mycorrhizal fungi, a presumably asexual plant root symbiont. *PLoS One* **8:** e80729.

Harumoto T, Anbutsu H, Fukatsu T (2014). Male-killing Spiroplasma induces sex-specific cell death via host apoptotic pathway. *PLoS Pathog* **10:** e1003956.

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM *et al* (1999). Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* **286:** 2165-2169.

Inagaki Y, Bessho Y, Osawa S (1993). Lack of peptide-release activity responding to codon UGA in Mycoplasma capricolum. *Nucleic Acids Res* **21:** 1335-1338.

Kai K, Furuyabu K, Tani A, Hayashi H (2012). Production of the quorum-sensing molecules N-acylhomoserine lactones by endobacteria associated with Mortierella alpina A-178. *Chembiochem* **13:** 1776-1784.

Kim JN, Roth A, Breaker RR (2007). Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. *Proc Natl Acad Sci U S A* **104:** 16092-16097.

Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J *et al* (2012). Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci U S A* **109:** E1173-1182.

Kloppholz S, Kuhn H, Requena N (2011). A secreted fungal effector of Glomus intraradices promotes symbiotic biotrophy. *Curr Biol* **21:** 1204-1209.

Kondorosi E, Mergaert P, Kereszt A (2013). A paradigm for endosymbiotic life: cell differentiation of Rhizobium bacteria provoked by host plant factors. *Annu Rev Microbiol* **67:** 611-628.

Ku C, Lo WS, Chen LL, Kuo CH (2013). Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in spiroplasma. *Genome Biol Evol* **5:** 1151-1164.

Lackner G, Moebius N, Hertweck C (2011a). Endofungal bacterium controls its host by an hrp type III secretion system. *ISME J* **5:** 252-261.

Lackner G, Moebius N, Partida-Martinez L, Hertweck C (2011b). Complete genome sequence of Burkholderia rhizoxinica, an Endosymbiont of Rhizopus microsporus. *J Bacteriol* **193:** 783-784.

Lackner G, Moebius N, Partida-Martinez LP, Boland S, Hertweck C (2011c). Evolution of an endofungal lifestyle: Deductions from the Burkholderia rhizoxinica genome. *BMC Genomics* **12:** 210.

Leclercq S, Giraud I, Cordaux R (2011). Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. *Mol Biol Evol* **28:** 685-697.

Ley DH, Berkhoff JE, McLaren JM (1996). Mycoplasma gallisepticum isolated from house finches (Carpodacus mexicanus) with conjunctivitis. *Avian Dis* **40:** 480-483.

Liu J, Maldonado-Mendoza I, Lopez-Meyer M, Cheung F, Town CD, Harrison MJ (2007). Arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. *Plant J* **50:** 529-544.

Lumini E, Bianciotto V, Jargeat P, Novero M, Salvioli A, Faccio A *et al* (2007). Presymbiotic growth and sporal morphology are affected in the arbuscular mycorrhizal fungus Gigaspora margarita cured of its endobacteria. *Cell Microbiol* **9:** 1716-1729.

Ma L, Jensen JS, Myers L, Burnett J, Welch M, Jia Q *et al* (2007). Mycoplasma genitalium: an efficient strategy to generate genetic variation from a minimal genome. *Mol Microbiol* **66:** 220-236.

Macdonald RM, Chandler MR (1981). Bacterium-like organelles in the vesicular-arbuscular mycorrhizal fungus Glomus caledonius. *New Phytol* **89:** 241-246.

McCutcheon JP, Moran NA (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* **2:** 708-718.

McCutcheon JP, von Dohlen CD (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* **21:** 1366-1372.

McCutcheon JP, Moran NA (2012). Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10:** 13-26.

McGowin CL, Anderson-Smits C (2011). Mycoplasma genitalium: an emerging cause of sexually transmitted disease in women. *PLoS Pathog* **7:** e1001324.

Melcher U, Sha Y, Ye F, Fletcher J (1999). Mechanisms of spiroplasma genome variation associated with SpV1-like viral DNA inferred from sequence comparisons. *Microb Comp Genomics* **4:** 29-46.

Miller RM, Reinhardt DR, Jastrow JD (1995). External Hyphal Production of Vesicular-Arbuscular Mycorrhizal Fungi in Pasture and Tallgrass Prairie Communities. *Oecologia* **103:** 17-23.

Mondo SJ, Toomer KH, Morton JB, Lekberg Y, Pawlowska TE (2012). Evolutionary stability in a 400-million-year-old heritable facultative mutualism. *Evolution* **66:** 2564-2576.

Morada M, Manzur M, Lam B, Tan C, Tachezy J, Rappelli P *et al* (2010). Arginine metabolism in Trichomonas vaginalis infected with Mycoplasma hominis. *Microbiology* **156:** 3734-3743.

Moran NA (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* **93:** 2873-2878.

Moran NA, McCutcheon JP, Nakabachi A (2008). Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42:** 165-190.

Morris JJ, Lenski RE, Zinser ER (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *MBio* **3**.

Mosse B (1970). Honey-coloured, sessile Endogone spores: II. Changes in fine structure during spore development. *Arch Mikrobiol* **74:** 129-145.

Mrazek J (2006). Analysis of distribution indicates diverse functions of simple sequence repeats in Mycoplasma genomes. *Mol Biol Evol* **23:** 1370-1385.

Nadal M, Paszkowski U (2013). Polyphony in the rhizosphere: presymbiotic communication in arbuscular mycorrhizal symbiosis. *Curr Opin Plant Biol* **16:** 473-479.

Naumann M, Schussler A, Bonfante P (2010). The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. *ISME J* **4:** 862-871.

Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JC, Andersson DI (2005). Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci U S A* **102:** 12112-12116.

Nunan LM, Lightner DV, Oduori MA, Gasparich GE (2005). Spiroplasma penaei sp. nov., associated with mortalities in Penaeus vannamei, Pacific white shrimp. *Int J Syst Evol Microbiol* **55:** 2317-2322.

Ochman H, Davalos LM (2006). The nature and dynamics of bacterial genomes. *Science* **311:** 1730-1733.

Olah B, Briere C, Becard G, Denarie J, Gough C (2005). Nod factors and a diffusible factor from arbuscular mycorrhizal fungi stimulate lateral root formation in Medicago truncatula via the DMI1/DMI2 signalling pathway. *Plant J* **44:** 195-207.

Paralanov V, Lu J, Duffy LB, Crabb DM, Shrivastava S, Methe BA *et al* (2012). Comparative genome analysis of 19 Ureaplasma urealyticum and Ureaplasma parvum strains. *BMC Microbiol* **12:** 88.

Parniske M (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Microbiol* **6:** 763-775.

Partida-Martinez LP, Hertweck C (2005). Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **437:** 884-888.

Partida-Martinez LP, Monajembashi S, Greulich KO, Hertweck C (2007). Endosymbiont-dependent host reproduction maintains bacterial-fungal mutualism. *Curr Biol* **17:** 773-777.

Pollack JD (2002). The necessity of combining genomic and enzymatic data to infer metabolic function and pathways in the smallest bacteria: amino acid, purine and pyrimidine metabolism in Mollicutes. *Front Biosci* **7:** d1762-1781.

Protsenko M (1975). Microorganism in the hyphae of mycorrhiza-forming fungus. *Mikrobiologiia* **44:** 1121-1124.

Qi J, Guo A, Cui P, Chen Y, Mustafa R, Ba X *et al* (2012). Comparative geno-plasticity analysis of Mycoplasma bovis HB0801 (Chinese isolate). *PLoS One* **7:** e38239.

Redecker D, Schussler A, Stockinger H, Sturmer SL, Morton JB, Walker C (2013). An evidence-based consensus for the classification of arbuscular mycorrhizal fungi (Glomeromycota). *Mycorrhiza* **23:** 515-531.

Remy W, Taylor TN, Hass H, Kerp H (1994). Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc Natl Acad Sci U S A* **91:** 11841-11843.

Rocha EP, Blanchard A (2002). Genomic repeats, genome plasticity and the dynamics of Mycoplasma evolution. *Nucleic Acids Res* **30:** 2031-2042.

Sato Y, Narisawa K, Tsuruta K, Umezu M, Nishizawa T, Tanaka K *et al* (2010). Detection of betaproteobacteria inside the mycelium of the fungus Mortierella elongata. *Microbes Environ* **25:** 321-324.

Schussler A, Walker C (2010). The *Glomeromycota:* a species list with new families and new genera.: Edinburgh & Kew, UK: The Royal Botanic Garden; Munich, Germany: Botanische Staatssammlung Munich; Oregon, USA: Oregon State University.

Shaw BM, Simmons WL, Dybvig K (2012). The Vsa shield of Mycoplasma pulmonis is antiphagocytic. *Infect Immun* **80:** 704-709.

Sirand-Pugnet P, Citti C, Barre A, Blanchard A (2007). Evolution of mollicutes: down a bumpy road with twists and turns. *Res Microbiol* **158:** 754-766.

Sloan DB, Moran NA (2013). The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biol Evol* **5:** 783-793.

Sturmer SL (2012). A history of the taxonomy and systematics of arbuscular mycorrhizal fungi belonging to the phylum Glomeromycota. *Mycorrhiza* **22:** 247-258.

Sugio A, Hogenhout SA (2012). The genome biology of phytoplasma: modulators of plants and insects. *Curr Opin Microbiol* **15:** 247-254.

Tardy F, Maigre L, Poumarat F, Citti C (2009). Identification and distribution of genetic markers in three closely related taxa of the Mycoplasma mycoides cluster: refining the relative position and boundaries of the Mycoplasma sp. bovine group 7 taxon (Mycoplasma leachii). *Microbiology* **155:** 3775-3787.

Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF *et al* (2012). Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proc Natl Acad Sci U S A* **109:** 8629-8634.

Tisserant E, Kohler A, Dozolme-Seddas P, Balestrini R, Benabdellah K, Colard A *et al* (2012). The transcriptome of the arbuscular mycorrhizal fungus Glomus intraradices (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. *New Phytol* **193:** 755-769.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R *et al* (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci U S A* **110:** 20117-20122.

Tokuriki N, Tawfik DS (2009). Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459:** 668-673.

Toomer KH, Chen X, Naito M, Mondo SJ, den Bakker HC, VanKuren NW *et al* (2014). Are the *Mycoplasma*-related endobacteria ancient parasites of arbuscular mycorrhizal fungi? *In Review.*

Tully JG, Whitcomb RF, Hackett KJ, Rose DL, Henegar RB, Bove JM *et al* (1994). Taxonomic descriptions of eight new non-sterol-requiring mollicutes assigned to the genus Mesoplasma. *Int J Syst Bacteriol* **44:** 685-693.

van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U *et al* (2003). Reductive genome evolution in Buchnera aphidicola. *Proc Natl Acad Sci U S A* **100:** 581-586.

Vancini RG, Benchimol M (2008). Entry and intracellular location of Mycoplasma hominis in Trichomonas vaginalis. *Arch Microbiol* **189:** 7-18.

Viscardi RM (2010). Ureaplasma species: role in diseases of prematurity. *Clin Perinatol* **37:** 393-409.

Volgmann T, Ohlinger R, Panzig B (2005). Ureaplasma urealyticum-harmless commensal or underestimated enemy of human reproduction? A review. *Arch Gynecol Obstet* **273:** 133-139.

von Dohlen CD, Kohler S, Alsop ST, McManus WR (2001). Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature* **412:** 433-436.

Vorburger C, Gehrer L, Rodriguez P (2010). A strain of the bacterial symbiont Regiella insecticola protects aphids against parasitoids. *Biol Lett* **6:** 109-111.

Waites KB, Katz B, Schelonka RL (2005). Mycoplasmas and ureaplasmas as neonatal pathogens. *Clin Microbiol Rev* **18:** 757-789.

Waites KB, Balish MF, Atkinson TP (2008). New insights into the pathogenesis and detection of Mycoplasma pneumoniae infections. *Future Microbiol* **3:** 635-648.

Wang B, Qiu YL (2006). Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16:** 299-363.

Wernegreen JJ (2012). Endosymbiosis. *Curr Biol* **22:** R555-561.

Williamson DL, Sakaguchi B, Hackett KJ, Whitcomb RF, Tully JG, Carle P *et al* (1999). Spiroplasma poulsonii sp. nov., a new species associated with male-lethality in Drosophila willistoni, a neotropical species of fruit fly. *Int J Syst Bacteriol* **49 Pt 2:** 611-618.

Ye F, Melcher U, Rascoe JE, Fletcher J (1996). Extensive chromosome aberrations in Spiroplasma citri Strain BR3. *Biochem Genet* **34:** 269-286.

Zheng X, Watson HL, Waites KB, Cassell GH (1992). Serotype diversity and antigen variation among invasive isolates of Ureaplasma urealyticum from neonates. *Infect Immun* **60:** 3472-3474.

Zheng Y, Anton BP, Roberts RJ, Kasif S (2005). Phylogenetic detection of conserved gene clusters in microbial genomes. *BMC Bioinformatics* **6:** 243.

CHAPTER 2

CANDIDATUS GLOMERIPLASMA MOENIUM, GEN. NOV., SP. NOV., AN ENDOBACTERIUM OF ARBUSCULAR MYCORRHIZAL FUNGI, REPRESENTING A NEW FAMILY OF MOLLICUTES, *GLOMERIPLASMATACEAE* FAM. NOV.[1]

2.1 - SUMMARY/ABSTRACT

Arbuscular mycorrhizal fungi (AMF, phylum *Glomeromycota*) are symbionts of most terrestrial plants. They commonly harbour endobacteria, referred to as MRE (*Mollicutes*/mycoplasma-related endobacteria). Phylogeny reconstructions based on 16S rRNA gene sequences cluster MRE with the representatives of the class *Mollicutes*, albeit without resolving their relationship with individual mollicute lineages. In this study, endobacteria associated with the AMF host *Rhizophagus clarus* are first representatives of MRE to be classified formally. Phylogenies based on amino acid sequences of 19 genes indicate that MRE form a discrete lineage sharing ancestry with the members of the family *Mycoplasmataceae*. Electron microscopy revealed that MRE are coccoid, ~0.5 µm in diameter, with a wall-like structure that is consistently present in all MRE, despite their phylogenetic placement in the *Mollicutes* class. MRE reside directly in the AMF cytoplasm and are uncultivable. On the basis of the unique lifestyle as endobacteria of AMF, the new genus and species *Candidatus* Glomeriplasma moenium gen. nov., sp. nov. is proposed for MRE of *R. clarus* NB112A, under the new family, *Glomeriplasmataceae* fam. nov. The type strain of *Ca.* Glomeriplasma moenium is associated with *R. clarus* NB112A.

[1]The results of this study will be submitted to the *International Journal of Systematic and Evolutionary Microbiology*, and is written according to their manuscript guidelines. All experiments except for TEM were performed by M.N. Only portions of the manuscript written by M.N. are displayed here.

2.2 - RESULTS/DISCUSSION

Arbuscular mycorrhizal fungi (AMF, phylum *Glomeromycota*) are obligate biotrophs forming symbiotic associations with the roots of most terrestrial plants (Gutjahr and Parniske 2013, Smith and Read 2008). They improve plant mineral nutrient uptake in exchange for photosynthates and are important members of terrestrial ecosystems. Based on electron microscopy studies, it has been known for decades that AMF harbour in their hyphae and spores bacterial endosymbionts of different cell morphologies, including coccoid cells (MacDonald and Chandler 1981, MacDonald et al 1982, Mosse 1970, Scannerini and Bonfante 1991). These coccoid endobacteria are unclassified and referred to as MRE (*Mollicutes*/mycoplasma-related endobacteria) based on the 16S rRNA gene phylogenies that cluster them with the members of the class *Mollicutes*, albeit without resolving their taxonomic position relative to individual mollicute lineages (Naumann et al 2010). MRE are found in the cytoplasm of nearly all lineages of AMF (Desirò et al 2013, Desirò et al 2014, Naumann et al 2010, Toomer et al 2014). The role of MRE in the biology of AMF is unknown. However, the molecular evolution patterns, including evidence of vertical transmission, abundant horizontal dispersal, and the lack of long-term codivergence with their hosts, are consistent with theoretical predictions and empirical data from bacterial antagonists of eukaryotes, suggesting that MRE may be parasites of *Glomeromycota* (Toomer et al 2014).

Endobacteria of *Rhizophagus clarus* NB112A

Spores of *Rhizophagus clarus* NB112A harbouring MRE were obtained from the International Culture Collection of Vesicular Arbuscular Mycorrhizal Fungi (INVAM), and maintained *in*

vitro in association with root-inducing T-DNA-transformed chicory roots on MSR medium (Cranenbrouck et al 2005) at 28ºC.

Fluorescent *in situ* hybridization (FISH) was performed on fixed and crushed AMF spores using the MRE-specific probe BLOgrBC (Naumann et al 2010) and the universal bacterial probe EUB338I (Amann et al 1990) as described by Naumann *et al.* (2010) with slight modifications, *i.e.* AMF spores were immobilized in polyacrylamide pads for the procedure, and probes were hybridized at a stringency of 30% formamide. Cells were visualized using the DeltaVision RT system (Applied Precision). Numerous MRE were detected in the cytoplasm of the *R. clarus* host (Figure 2.1). As a second step, spores of *R. clarus* were subjected to high-pressure/freeze-fixation in order to better preserve fungal and bacterial cytology, and observed under transmission electron microscope (Desirò et al 2014). Figure 2.2A illustrates a portion of the fungal protoplasm where a fungal nucleus is close to an endobacterium residing directly in the fungal cytoplasm without any surrounding membrane. The bacterium is coccoid in shape, as previously described in MRE of the *Gigaspora margarita* host (Desirò et al 2014). However, other *R. clarus*-associated MRE may have a slightly different shape when, for example, they are compressed between the lipid bodies (not shown). Figure 2.2B provides details of bacterial morphology; a homogenous electron-dense layer is consistently present outside the membrane, while many ribosomes populate the cytoplasm.

PCR amplification of MRE 16S rRNA gene sequences was performed using the primers 109F1, 109F2, 1184R1, 1184R2, and 1184R3 (Naumann et al 2010). Amplification was carried out on 3 randomly selected *R. clarus* spores, using Phusion® Hot Start Flex DNA Polymerase (New England BioLabs), following the manufacturer's directions, with initial denaturation of 3 min at 98ºC, followed by 20 cycles of 10 sec at 98ºC, 10 sec at 60ºC, and 1 min at 72ºC.

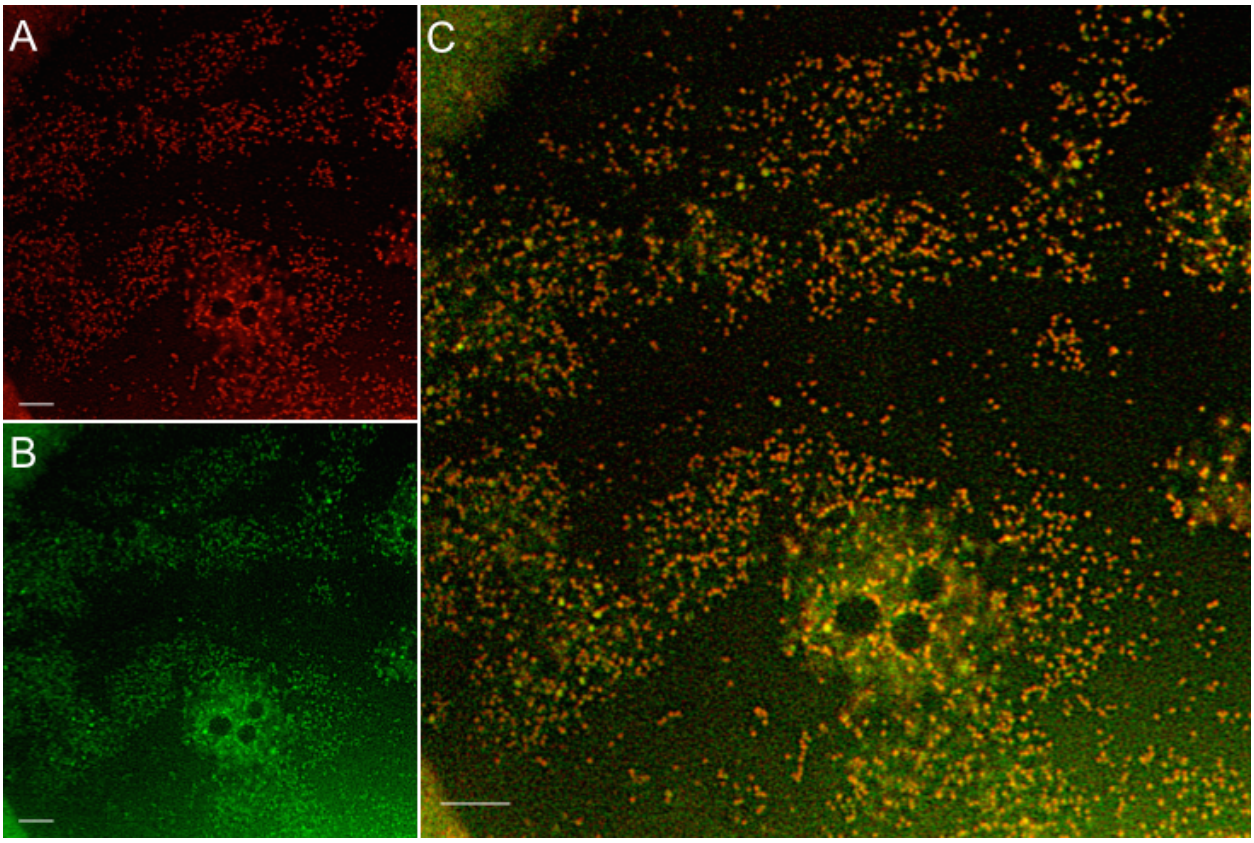


Figure 2.1. FISH of MRE within the cytoplasm of a crushed spore of the AMF *Rhizophagus clarus*. **A**. MRE visualized with the MRE-specific probe, BLOgrBC (red). **B**. MRE visualized with the universal bacterial probe EUB338I (green). **C**. An overlay of A and B. Scale bars, 5 μm.

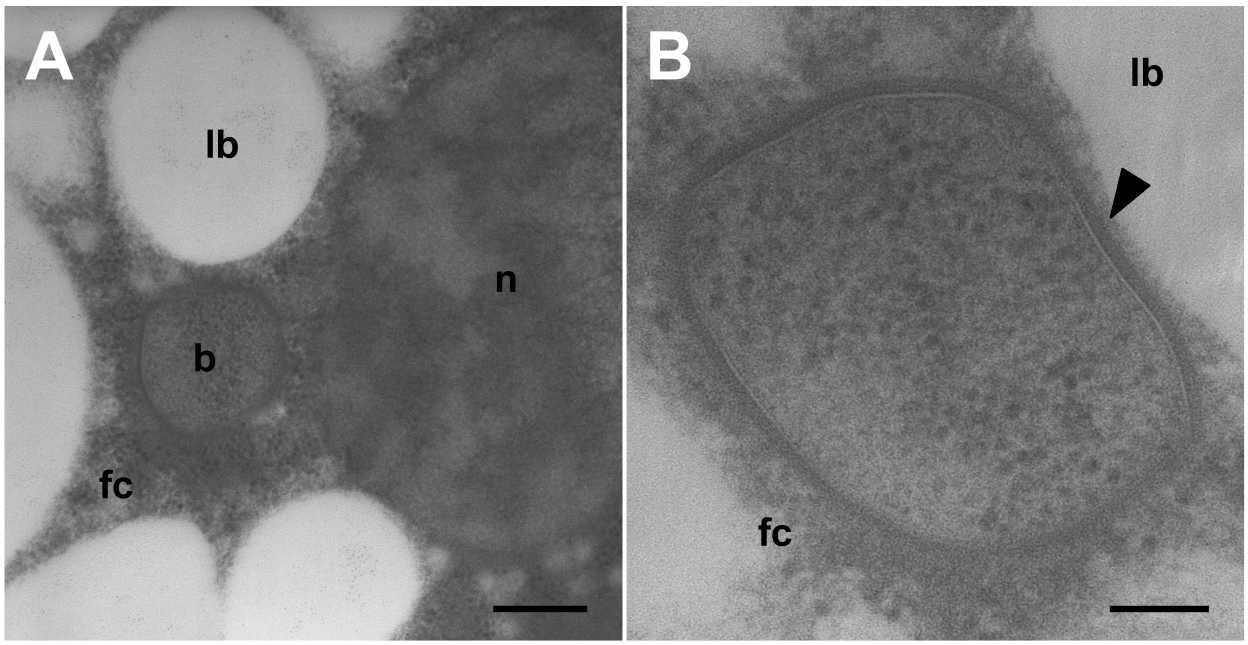


Figure 2.2. Transmission electron micrographs of MRE in the cytoplasm of *Rhizophagus clarus* NB112A. **A.** Endobacteria are directly embedded in the fungal cytoplasm (fc), without any surrounding membrane. **B.** A homogenous electron-dense layer (arrowhead) is consistently present outside the membrane of the endobacteria, while many ribosomes populate their cytoplasm. Scale bars: A, 0.32 µm; B, 0.10 µm. lb = lipid bodies; n = fungal nucleus; b = MRE bacteria. This image was provided courtesy of Alessandro Desirò, from the Bonfante Group at the University of Turin.

Amplicons were cloned into CloneJET pJET1.2 vector (Thermo Scientific) and transformed into One Shot® Top10 chemically competent *E. coli* cells (Invitrogen). PCR was performed on individual colonies to amplify their pJET1.2 vector inserts, following the manufacturer's protocol, and amplicons were sequenced. To examine the taxonomic relationship of the MRE in *R. clarus* with MRE in other AMF host species, 16S rRNA sequences of MRE in the AMF hosts *Racocetra verrucosa* VA103A (obtained from INVAM) and *Claroideoglomus etunicatum* CA-OT135 (VanKuren et al 2013) were also generated as above. For phylogenetic analyses, two representative 16S rRNA sequences from each spore (6 MRE sequences per AMF host species) were used. Non-MRE sequences were obtained from the Integrated Microbial Genomes, IMG (Markowitz et al 2012). Nucleotide sequences were aligned in MUSCLE (Edgar 2004) and phylogenies reconstructed using the GTR+I+Γ nucleotide substitution model implemented in PhyML (Guindon et al 2010) with 1,000 bootstrap replications, and in MrBayes v.3.1.2 (Ronquist and Huelsenbeck 2003) with analyses conducted for 2,000,000 generations with 25% burn-in. While the MRE 16S rRNA gene sequences clustered together forming a well-supported clade, the relationship of MRE with other lineages of the Mollicutes remained unresolved (Figure 2.3). The sequences of MRE from the three different hosts showed different levels of diversity: 99.9% similarity in *R. clarus*, 97% similarity in *R. verrucosa*, and 88.3% similarity in *C. etunicatum*, consistent with previously reported patterns of MRE 16S rRNA gene sequence diversity and divergence within host individuals (Desirò et al 2014, Naumann et al 2010, Toomer et al 2014). Furthermore, MRE from all three AMF species were divergent from each other, with an average of 87% similarity. Finally, MRE from *R. clarus* formed a well-supported clade within the MRE lineage with all 16S rRNA sequences being virtually identical to each other.

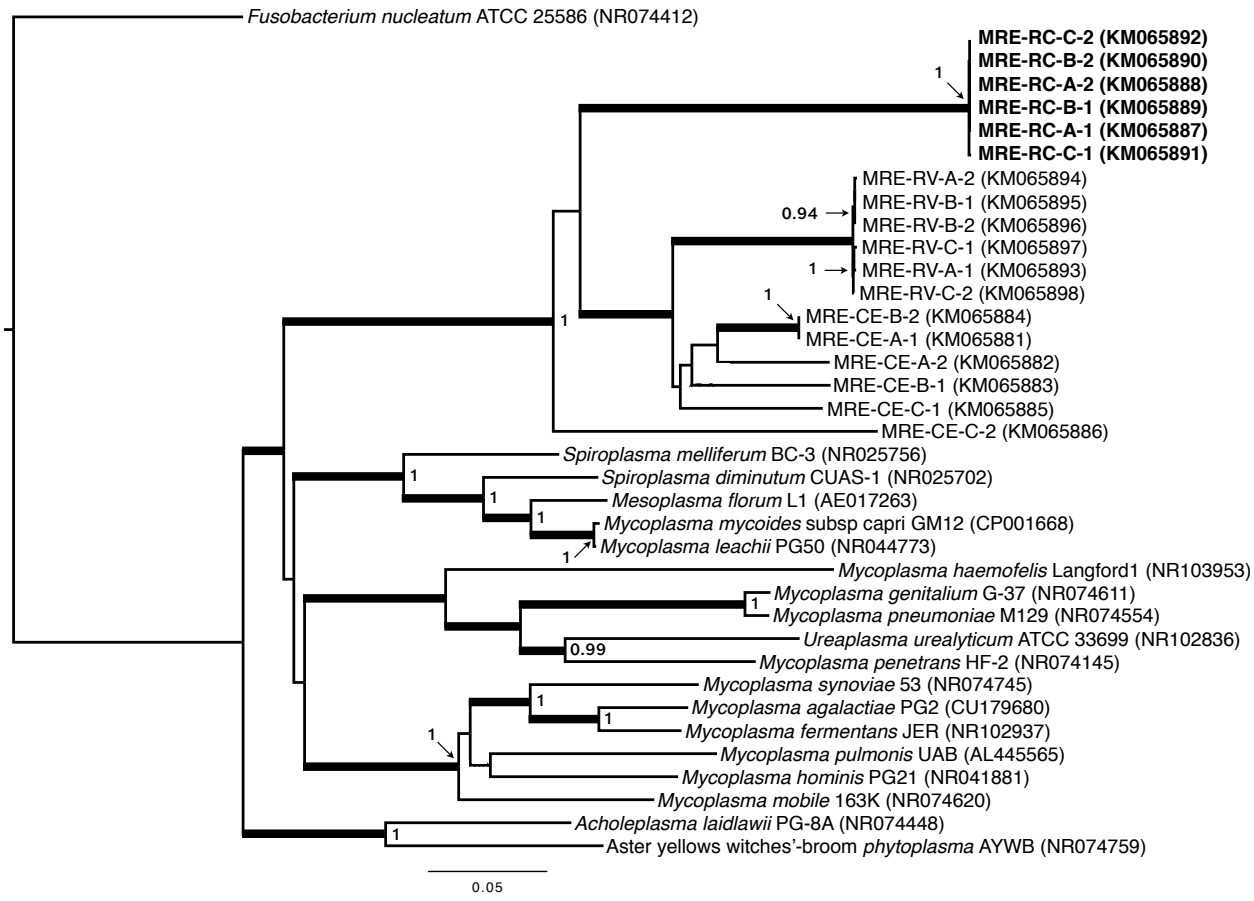


Figure 2.3. Phylogenetic position of MRE based on 16S rRNA gene sequences. Bayesian posterior probabilities greater than 0.90 are indicated. Branches with Maximum Likelihood bootstrap support greater than 70% are thickened. MRE-RC indicates MRE from *R. clarus* (bold text); MRE-RV, MRE from *R. verrucosa*; MRE-CE, MRE from *C. etunicatum*. Letters following "MRE-XX" are spore designations, and the subsequent numbers are individual sequence designations.

Because of the high intrahost similarity of 16S rRNA gene sequences, we selected MRE from *R. clarus* to be classified taxonomically.

To resolve the relationship between MRE of *R. clarus* and other lineages within the Mollicutes class, we conducted a multi-gene phylogenetic reconstruction using amino acid sequences of 19 conserved genes (*dnaG, infC, nusA, rplA, rplB, rplC, rplE, rplF, rplM, rplN, rplP, rplT, rpmA, rpsB, rpsC, rpsE, rpsJ, rpsS, smpB*), selected based on the Genomic Encyclopaedia of Bacteria and Archaea (Wu et al 2009). Sequences of these genes were extracted from the *de novo* sequenced metagenomes of MRE associated with *R. clarus* and *R. verrucosa* (Chapter 3). Sequences for non-MRE species were obtained from IMG (Markowitz et al 2012). Amino acid sequences were concatenated and aligned using MUSCLE (Edgar 2004). Maximum Likelihood analysis was conducted using PhyML (Guindon et al 2010) under the WAG model run with 1,000 bootstrap replications. Bayesian analyses were performed using MrBayes v.3.1.2 (Ronquist and Huelsenbeck 2003) with the mixed amino acid+I+Γ model run for 1,000,000 generations with 25% burn-in. These reconstructions indicated that MRE form a distinct lineage sharing ancestry with the representatives of the family *Mycoplasmataceae* (Figure 2.4).

Cultivation of MRE was attempted, using approaches consistent with their taxonomic classification, endobacterial lifestyle, and metabolic capabilities inferred from the draft genome (Chapter 3). Every possible combination of the following four factors was tested: (i) cultivation medium, including Brain Heart Infusion, BHI (Bacto), PPLO Broth Base (BBL), 2x BHI, 2x PPLO, and agar Noble (Difco) as solidifier, (ii) supplement, including AMF spore extract, horse serum (Sigma), bovine serum (Sigma), and porcine serum (Sigma) with all sera at 1 to 20% in 5% increments, yeast extract and TC yeastolate (Bacto) at 0.1%, 0.25%, 0.5% and 1%, and

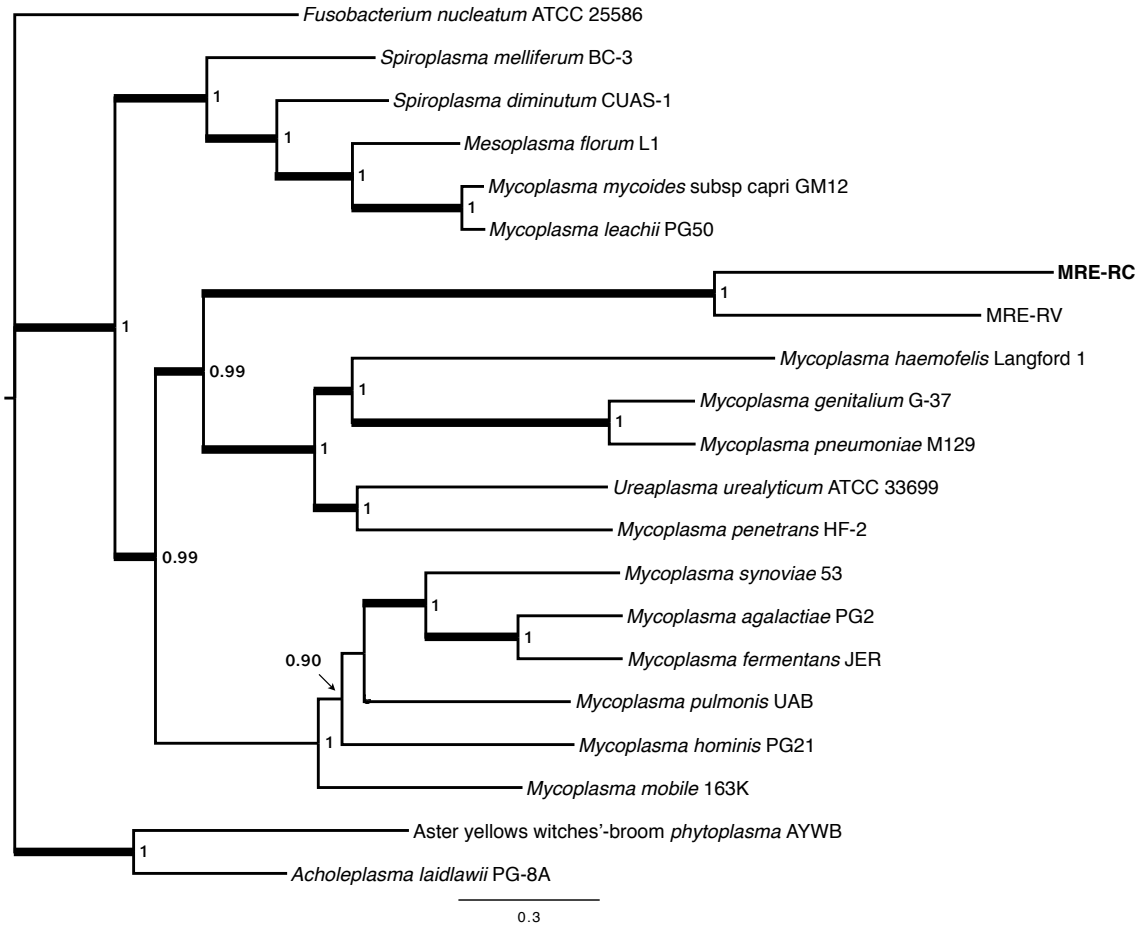


Figure 2.4. Phylogenetic position of MRE based on concatenated amino acid sequences of 19 conserved proteins. Bayesian posterior probabilities greater than 0.90 are indicated. Branches with Maximum Likelihood bootstrap supports greater than 70% are thickened. MRE-RC indicates MRE from *R. clarus* (bold text); MRE-RV, MRE from *R. verrucosa*.

Tween®80 at 0.05% and 0.5% (Sigma), (iii) temperature, including ambient temperature, 28ºC, 30ºC, and (iv) atmosphere, including ambient, microaerophilic, increased CO_2, and anaerobic atmosphere. Each medium and supplement condition was prepared as a liquid culture and inoculated at day 0 with AMF filtrate containing MRE cells, followed by incubation at every combination of temperature and atmospheric conditions. On day 0, 1, 3, 7, 14, 21, and 30, a portion of the liquid culture was sub-cultured onto a solid medium of the same type, and incubated for further 14 days, in the same temperature and atmospheric conditions as before. Any colonies that arose were genotyped by 16S rRNA gene sequencing, but none was identified as MRE.

Based on phylogenetic evidence, MRE form a novel clade in the class *Mollicutes*, sharing common ancestry with the members of the family *Mycoplasmataceae*. Biologically, MRE share with other *Mycoplasmataceae* a lifestyle of metabolic dependence on the eukaryotic host. However, unlike other *Mycoplasmataceae*, MRE are associated with fungal hosts in the phylum *Glomeromycota*, and reside directly in the host cytoplasm. To recognize these distinctions, a new family is proposed in the *Mycoplasmatales*, the *Glomeriplasmataceae* fam. nov. This family currently includes one new genus, *Glomeriplasma* gen. nov., and one new species *Glomeriplasma moenium* sp. nov. found associated with the AMF host *R. clarus* NB112A. Cultivation experiments suggest that the *Glomeriplasma* endobacteria are uncultivable, and, consequently, are given a *Candidatus* designation (Murray and Stackebrandt 1995). The endobacteria of the AMF *R. clarus*, are, therefore, *Candidatus* Glomeriplasma moenium, gen nov., sp. nov., of the family *Glomeriplasmataceae*, fam. nov.

Description of *Candidatus* Glomeriplasma gen. nov.

Glomeriplasma (Glo.me.ri.pla'sma, from *glomeris*, Latin, genitive of *glomus*, a ball-shaped mass, referring to the name of the *Glomeromycota* host fungi, and *plasma*, Greek, that which is molded/shaped).

Cells are coccoid, 0.46 – 0.61 μm in diameter (measured in 8 cells), and found in the cytoplasm of *Glomeromycota*. Based on electron micrographs, an electron dense, wall-like substance surrounds the cell, an unusual feature for the wall-less *Mollicutes* class. Draft genomes of MRE-RC and MRE-RV suggest that this substance is not a Gram-positive cell wall (Chapter 3). The *Glomeriplasma*-specific primers, targeting their 16S rRNA gene sequence are available as 109F and 1184R, as well as *Glomeriplasma*-specific probes BLOgrBC (Naumann et al 2010) and BLOsADf2 (Desirò et al 2013) for FISH studies in the AMF host cytoplasm.

Description of *Candidatus* Glomeriplasma moenium gen. nov., sp. nov.

Glomeriplasma moenium (*Glomeriplasma,* as above; *moenium,* moe'nium, genitive of *moenia*, Latin, of walls/fortifications produced by the distributed efforts of many).

Representatives of this species include uncultivable endobacteria of *R. clarus* NB112A. [(*Mollicutes*) NC; NA; C; NAS (GenBank nos KM065881-KM065898); S (*Rhizophagus clarus* NB112A, cytoplasm of spores and mycelium)]. The type strain of *Ca.* Glomeriplasma moenium is associated with *R. clarus* NB112A.

Description of *Glomeriplasmataceae* fam. nov.

Glomeriplasmataceae (*Glomeriplasma*, as above; *-aceae, family suffix*).

The 16S rRNA gene phylogeny indicates that MRE form a distinct and well supported clade within the Mollicutes (Figure 2.3). The multi-gene phylogeny suggests that while *Ca.* Glomeriplasma shares ancestry with the members of the *Mycoplasmataceae* family, it represents a discrete lineage associated with fungi rather than with animals (Figure 2.4). To recognize these peculiarities, the new family *Glomeriplasmataceae* is proposed. The type and only genus is *Ca.* Glomeriplasma.

2.3 - ACKNOWLEDGEMENTS

I would like to thank our collaborators Alessandro Desirò and Paola Bonfante, University of Turin, for their electron microscopy images and for discussions on selecting a name for the MRE. We thank J. Morton for fungal isolates from the INVAM collection, A. Kaech, University of Zurich, Switzerland, and A. Faccio, ISPP, CNR-Turin, for the high-pressure cryo preparation. This work was supported by the National Science Foundation grant DEB-0918880 and by the Cornell University Agricultural Experiment Station federal formula funds, Project NYC-153434, received from the National Institutes for Food and Agriculture, U.S. Department of Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the U.S. Department of Agriculture.

REFERENCES

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA (1990). Combination of 16S ribosomal RNA-targeted oligonucleotide probes with flow-cytometry for analyzing mixed microbial-populations. *Applied and Environmental Microbiology* **56:** 1919-1925.

Cranenbrouck S, Voets L, Bivort C, Renard L, Strullu DG, Declerck S (2005). Methodologies for *in vitro* cultivation of arbuscular mycorrhizal fungi with root organs. In: Declerck S, Strullu DG, Fortin A (eds). *In Vitro Culture of Mycorrhizas*. Springer-Verlag: Berlin, Heidelberg. pp 341-375.

Desirò A, Naumann M, Epis S, Novero M, Bandi C, Genre A *et al* (2013). *Mollicutes*-related endobacteria thrive inside liverwort-associated arbuscular mycorrhizal fungi. *Environmental Microbiology* **15:** 822-836.

Desirò A, Salvioli A, Ngonkeu EL, Mondo SJ, Epis S, Faccio A *et al* (2014). Detection of a novel intracellular microbiome hosted in arbuscular mycorrhizal fungi *ISME Journal* **8:** 257–270.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792-1797.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* **59:** 307-321.

Gutjahr C, Parniske M (2013). Cell and developmental biology of arbuscular mycorrhiza symbiosis. *Annual Review of Cell and Developmental Biology* **29:** 593-617.

MacDonald RM, Chandler MR (1981). Bacterium-like organelles in the vesicular-arbuscular mycorrhizal fungus *Glomus caledonius*. *New Phytologist* **89:** 241-246.

MacDonald RM, Chandler MR, Mosse B (1982). The occurrence of bacterium-like organelles in vesicular-arbuscular mycorrhizal fungi. *New Phytologist* **90:** 659-663.

Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y *et al* (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research* **40:** D115-D122.

Mosse B (1970). Honey-coloured, sessile *Endogone* spores: II. Changes in fine structure during spore development. *Archiv fur Mikrobiologie* **74:** 129.

Murray RG, Stackebrandt E (1995). Taxonomic note: implementation of the provisional status Candidatus for incompletely described procaryotes. *Int J Syst Bacteriol* **45:** 186-187.

Naumann M, Schüßler A, Bonfante P (2010). The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. *ISME Journal* **4:** 862-871.

Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572-1574.

Scannerini S, Bonfante P (1991). Bacteria and bacteria like objects in endomycorrhizal fungi (Glomaceae). In: Margulis L, Fester R (eds). *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*. MIT Press: Cambridge, MA. pp 273-287.

Smith SE, Read DJ (2008). *Mycorrhizal Symbiosis*, Third edn. Academic Press: New York, NY.

Toomer KH, Chen X, Naito M, Mondo SJ, den Bakker HC, VanKuren NW *et al* (2014). Are the mycoplasma-related endobacteria ancient parasites of arbuscular mycorrhizal fungi? *In review*.

VanKuren NW, den Bakker HC, Morton JB, Pawlowska TE (2013). Effective population size, rRNA gene diversity, and evolutionary longevity of asexual Glomeromycota. *Evolution* **67:** 207–224.

Wu DY, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN *et al* (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462:** 1056-1060.

CHAPTER 3

THE METAGENOME SEQUENCING OF THE MYCOPLASMA-RELATED ENDOBACTERIA OF ARBUSCULAR MYCORRHIZAL FUNGI REVEALS HIGHLY PLASTIC MINIMAL GENOMES WITH GENES OF FUNGAL ORIGIN AND ADAPTATIONS TO LIFE IN THE FUNGAL CYTOPLASM[1]

3.1 - ABSTRACT

Arbuscular mycorrhizal fungi (AMF, phylum Glomeromycota) are obligate biotrophs that form symbiotic associations with the roots of the majority of terrestrial plants. They provide essential minerals to their plant hosts and receive photosynthates in return. All major lineages of AMF harbour endobacteria classified as Mollicutes and referred to as the Mollicutes/mycoplasma-related endobacteria (MRE). Except for the molecular evolution patterns suggesting that MRE may be parasites of AMF, virtually nothing is known about the life history of these endobacteria. To understand MRE biology, we sequenced metagenomes of MRE populations associated with three AMF host species. Each AMF species harboured a genetically distinct group of MRE, with some MRE populations displaying unusually high levels of diversity within the same host. In addition, all three MRE populations showed extensive chromosomal rearrangements, which we attribute to genomic recombination as well as to activity of mobile elements and a history of plectroviral invasion. The MRE genomes are characterized by a highly reduced gene content indicating metabolic dependence on the fungal host, with the mechanism of energy production remaining unclear. Several MRE genes encode proteins with domains involved in protein-

[1]The results of this study will be submitted to *PLoS Biology*, and is written according to their manuscript guidelines. All experiments were performed by and manuscript written by M.N.

protein interactions with eukaryotic hosts. In addition, the MRE genomes harbour genes horizontally acquired from AMF, including SUMO proteases specific to the SUMOylation systems of eukaryotes, which the MRE likely use to manipulate their fungal host. The extent of MRE genome plasticity and reduction, along with the large number of horizontally acquired host genes, suggest a highly coevolved antagonistic interaction between the partners. Together with the ubiquity of the MRE-Glomeromycota associations, these features emphasize the significance of MRE in the Glomeromycota biology.

3.2 - AUTHOR SUMMARY

Arbuscular mycorrhizal fungi are soil fungi that live in association with the majority of land plants worldwide. They supply plants with mineral nutrients in exchange for photosynthates. These fungi have recently been discovered to harbour endobacteria classified in the Mollicutes class. The Mollicutes are a unique group of bacteria that have undergone reductive genome evolution, and all members form intimate associations with eukaryotic hosts, either as antagonists or rare mutualists. In this study, we explored metagenomes of three populations of the endobacteria in three separate arbuscular mycorrhizal fungal species. We discovered that, while the endobacteria are metabolically dependent on their fungal hosts, they have uniquely plastic genomes, and the losses of important genes are prevented through a mechanism involving chromosomal rearrangements and recombination of rearranged genomes. We also found that the endobacterial genomes encode a multitude of genes horizontally transferred from the fungal host, and that many of these gene products likely interact with fungal host proteins. Overall, the endobacterial genomes reveal a tightly knit network of potentially antagonistic interactions with

the fungal host, ensuring evolutionary and ecological stability of the association. This novel symbiosis highlights the importance of endobacteria in eukaryotic systems.

3.3 - INTRODUCTION

Arbuscular mycorrhizal fungi (AMF) of the phylum Glomeromycota are obligate biotrophs that form symbiotic associations with the roots of the majority of terrestrial plant species (Wang and Qiu 2006). AMF provide essential mineral nutrients from the soil to the plant host and receive up to 20% of the plant's photosynthetic energy in return (Bago et al 2000, Fellbaum et al 2012, Parniske 2008). AMF are emerging as the key constituents of the future sustainable agriculture, with the goal of optimizing mycorrhizal associations of crop plants to replace the use of non-renewable mineral fertilizers (Gianinazzi et al 2010).

For almost half a century, AMF have been known to harbour morphologically diverse bacteria in their hyphae and spores (Macdonald and Chandler 1981, Mosse 1970). Recently, one of these bacterial morphotypes has been identified as a novel member of the class Mollicutes (Naumann et al 2010), and is referred to as the Mollicutes/mycoplasma-related endobacteria (MRE) (Naumann et al 2010, Toomer et al 2014). Mollicutes are a unique class of bacteria, characterized by reductive evolution and minimal genomes. Consequently, all members of Mollicutes are intimate associates of eukaryote hosts (Sirand-Pugnet et al 2007a), acting as parasites or rare mutualists (Jaenike et al 2010). MRE are the first members of Mollicutes to be found associated with fungi. They reside directly in the host cytoplasm and have been detected in all major lineages of AMF worldwide (Bianciotto et al 2003, Naumann et al 2010). Their molecular evolution patterns, including evidence of predominantly vertical transmission,

occasional horizontal dispersal, and the lack of long-term codivergence with Glomeromycota, suggest that MRE may be parasites of AMF (Toomer et al 2014).

To generate direct insights into the MRE biology, we sequenced metagenomes of MRE associated with three divergent AMF species. Our aim was to determine metabolic capabilities of MRE and infer their relationship with the AMF hosts. Furthermore, we were interested in assessing whether MRE associated with divergent hosts followed similar trajectories of genome evolution. While the MRE gene content suggests that they are metabolically dependent on the AMF hosts, their genomes are highly plastic, with genome rearrangements facilitated by recombination machinery and transposable elements. The MRE genomes encode proteins with domains involved in protein-protein interactions in eukaryotes. They also contain products of inter-domain horizontal transfer from the fungal host. Overall, the MRE genomes show a high degree of potentially antagonistic coevolution with their fungal hosts, indicating the significance of these endobacteria in the biology of Glomeromycota.

3.4 - RESULTS

MRE diversity, metagenome sequencing, and analysis

Previous surveys revealed that MRE are abundant in all major lineages of Glomeromycota worldwide (Toomer et al 2014). To explore MRE genomes across AMF host diversity, we selected three AMF species: *Claroideoglomus etunicatum* (CE, family Claroideoglomeraceae), *Racocetra verrucosa* (RV, Gigasporaceae), and *Rhizophagus clarus* (RC, Glomeraceae). Unlike other vertically transmitted endosymbionts that display genetic homogeneity, facilitating assembly of their genomes (Gardebrecht et al 2012, Tamas et al 2002), MRE have been shown to be diverse within host individuals (Desiro et al 2014, Naumann et al 2010, Toomer et al 2014).

Consequently, to better understand the population structure of MRE within each focal host, we conducted a phylogenetic analysis of cloned MRE 16S rRNA gene sequences obtained from individual AMF spores (Figure 3.1). While MRE sequences from all three hosts clustered together, forming a well-supported MRE clade, each host harboured a distinct population with diversity ranging from 99.9% similarity in RC through 97% similarity in RV to 88.3% similarity in CE.

As MRE are uncultivable (Chapter 2), their cells were extracted from multiple spores of the respective AMF hosts and metagenomes sequenced using the Illumina HiSeq platform with 2 × 100 paired-end reads, followed by quality control, assembly, and annotation. The statistics of the metagenomic assemblies for each MRE population are presented in Table 3.1. As expected based on 16S rRNA gene sequence diversity, MRE contigs revealed genomic complexity, with chromosomal rearrangements apparent through alterations of gene synteny. Although this biological feature of MRE prevents the assembly of a single genome from any of the focal populations, biological capabilities of the MRE population as a whole within their AMF hosts can be assessed. In particular, the metagenomic assemblies show that the MRE genomes have a low GC content, consistent with other mollicute species (Moran et al 2008, Thompson et al 2011). Furthermore, based on the number of contigs that contain the same gene content, we established that the MRE genomes are fairly small, with a genome size of less than 700,000 bp.

Plasticity of MRE genomes

Despite the 16S rRNA gene phylogeny indicating homogeneity of the MRE-RC population, the MRE-RC contigs reveal vast chromosomal rearrangements, with stretches of DNA sequences with 100% identity between 2 contigs followed by stretches of DNA entirely divergent from

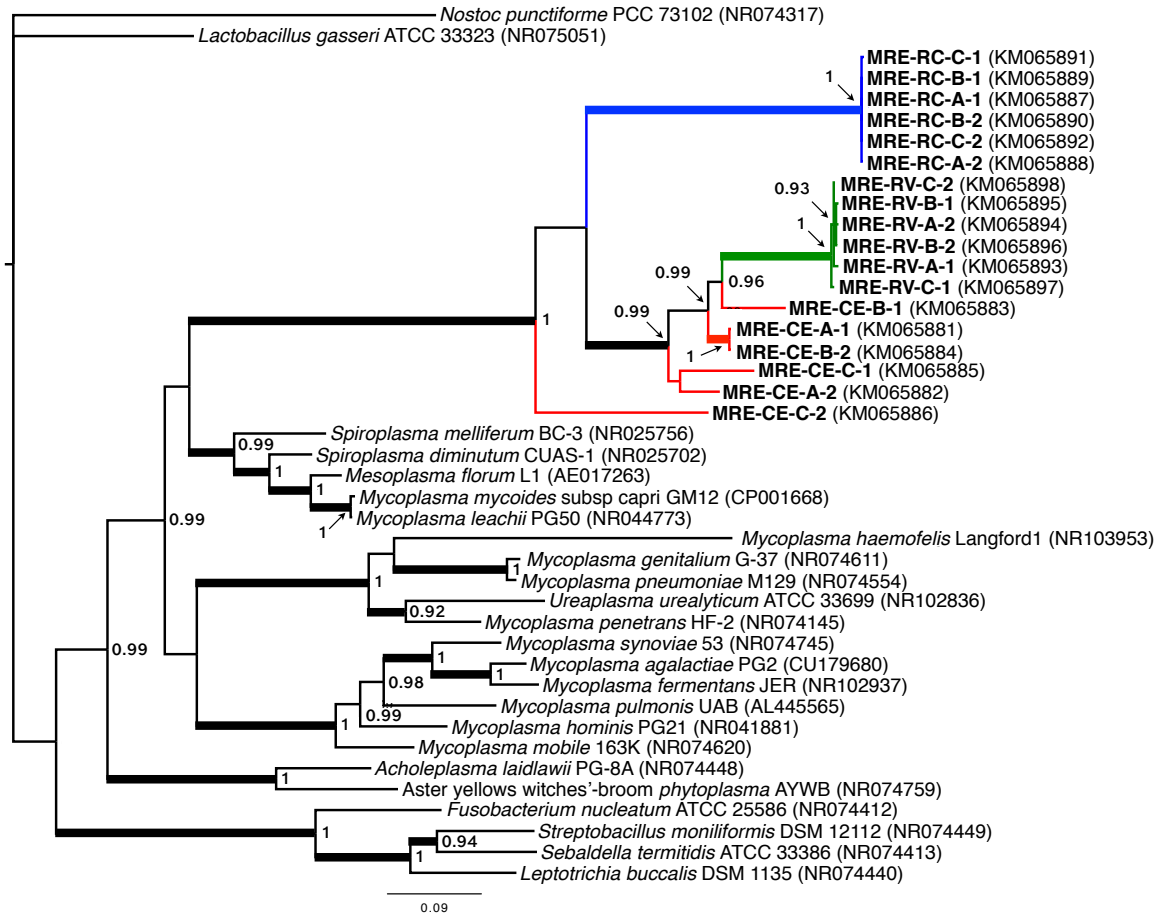


Figure 3.1. Phylogeny of MRE associated with three AMF hosts reveals inter- and intra-population diversity. Cloned 16S rRNA sequences were obtained from three individual spores (A, B, or C) representing population MRE-RC, MRE-RV, and MRE-CE (bolded). Numbers following spore designations indicate individual sequences. Bayesian posterior probabilities greater that 0.90 are shown at nodes. Branches with Maximum Likelihood bootstrap support over 70% are thickened.

Table 3.1. Metagenome assembly statistics of MRE associated with three AMF hosts. Numbers in parenthesis represent duplicate contigs.

	MRE-RC	MRE-RV	MRE-CE
Number of chromosomal contigs	34	99	67
Combined size of contigs (bp)	739,936 (125,736)	1,227,948 (70,737)	662,952 (55,942)
GC content (%)	32.0	33.6	34.3
No. of CDS	1034 (180)	1855 (112)	1078 (88)
No. of orphan genes	640	1316	732
tRNA	24	29	34

each other. Analysis of the junctions between regions of sequence similarity and dissimilarity revealed two main sources of the chromosomal rearrangements in the MRE population of RC: mobile element activity and recombination (Supplementary Table 3.1). These two forces are also important in MRE populations associated with RV and CE. MRE genomes contain an unusually high number of transposases: 31 of the 394 non-orphan genes are transposases in MRE-RC, 19 of the 539 in MRE-RV, and 5 out of 346 in MRE-CE. Most of them are located at the junctions where chromosomal rearrangements are evident. However, some junctions lack noticeable gene products that may have caused rearrangements, and we believe this to be a consequence of recombination events that occur in the genome. Specifically, the MRE genomes in all three populations are riddled with simple sequence repeats (SSRs), genomic features common in many *Mycoplasma* species (Mrazek 2006). These sequence features are known to facilitate homologous recombination events in the *Mycoplasma* genomes, and create high intraspecies variability (Sirand-Pugnet et al 2007b, Sogaard et al 2002). Furthermore, genes involved in DNA recombination are present in the MRE genomes (Supplementary Table 3.2), a feature shared with other *Mycoplasma* genomes (Moran et al 2008). In particular, MRE genomic assemblies display an overabundance of the XerC/D tyrosine recombinase genes (Supplementary Table 3.2). The Xer recombination system is normally used by bacteria to maintain monomeric state of chromosomes (dimer resolution), but is also responsible for causing inversions, and integration and excision of gene cassettes (Das et al 2013). The Xer recombinase system has been implicated as the source of DNA rearrangements in other *Mycoplasma* species (Ron et al 2002). The unusually high copy number of the XerC/D genes, combined with the AT-rich genomes of MRE, increase the likelihood of a Xer recombinase target sites, and suggests that the Xer recombinase system is likely important in the chromosome structure and plasticity of the

MRE genomes. In fact, previous analyses of MRE 16S rRNA sequences also found evidence of recombination in MRE populations (Desiro et al 2014, Toomer et al 2014).

In addition to mobile elements and recombination events as factors responsible for genome plasticity (Supplementary Table 3.1), the genomes MRE-RV and MRE-CE display evidence of a previous plectroviral invasion, a phage known to infect *Spiroplasma* species (Carle et al 2010, Ku et al 2013). Though the MRE genomes harbour only partial plectroviral genes, some of the genomic rearrangements may be due to the past activity of this phage, as seen in the *Spiroplasma* genomes, making their assemblies also difficult.

Limited metabolic capabilities of MRE

MRE biological functionalities are displayed in Figure 3.2. The pathways are based on KEGG Orthology (KO) assignments, and proteins not assigned KO identifiers are not included. As expected of the representatives of Mollicutes with intracellular lifestyle, the genomic capacities of MRE are very limited, and the majority of their functional capabilities are that of basic cell maintenance. Even the most essential cellular functions have been reduced to a minimum. For instance, all 3 MRE populations have lost all but the α subunit of the DNA polymerase III holoenzyme, and their RNA polymerase consists only of α, β, and β' subunits. MRE are incapable of amino acid and nucleic acid biosynthesis, and these metabolites must be obtained from the AMF host cytoplasm. Furthermore, as with other *Mycoplasma* species, the MRE genomes do not encode for enzymatic capabilities responsible for the TCA cycle and oxidative phosphorylation. Consequently, the mechanism of energy production in MRE remains unclear.

Electron micrographs depicting MRE within the cytoplasm of various AMF fungi suggest that MRE are surrounded by a wall-like substance, unexpected for bacteria of the wall-less

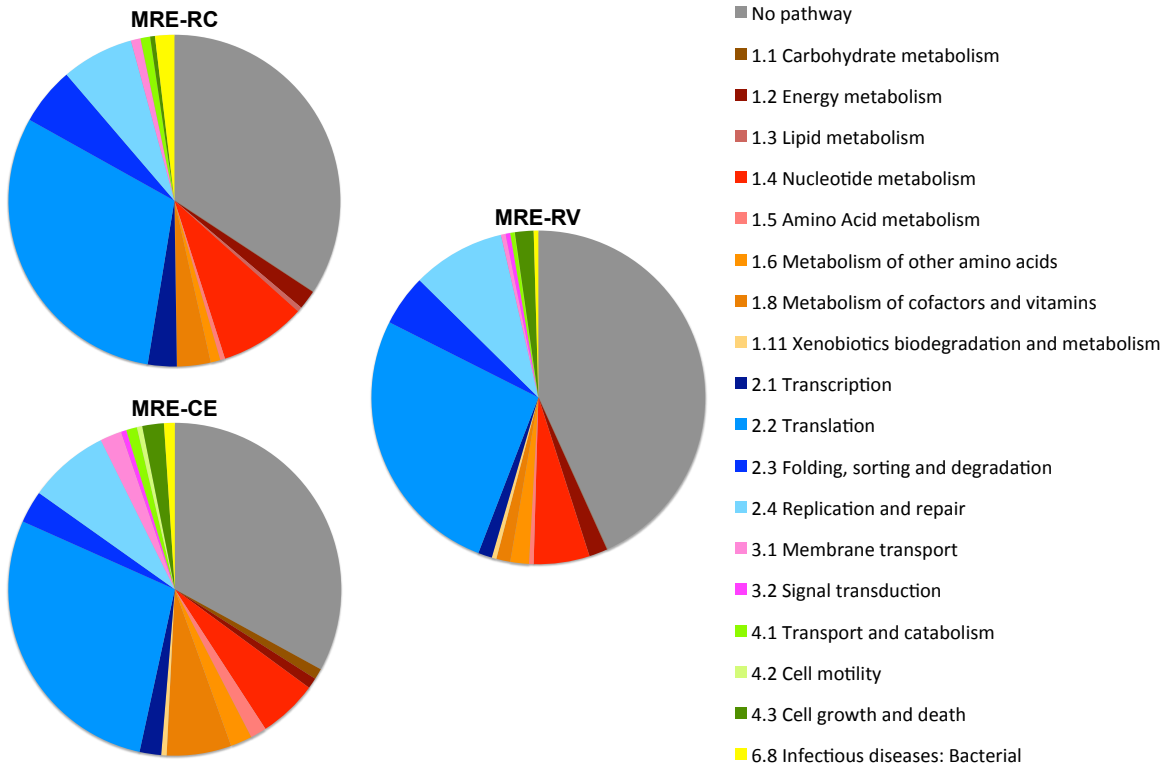


Figure 3.2. Functional capabilities of MRE identified by KEGG Orthology (KO) assignments and their corresponding KEGG pathways. The protein sequences of the three MRE populations were assigned KO identifiers for functional analysis. The main functional categories involved are 1.0 Metabolism, 2.0 Genetic Information Processing, 3.0 Environmental Information Processing, 4.0 Cellular Processes, and 6.0 Human Diseases. The subcategories are indicated in the colour legend of the figure. Proteins that were assigned KO identifiers, but are not part of any KEGG pathway, are labelled as "No pathway". MRE proteins with no KO identifiers assigned are not included in the figure.

Mollicutes class (Desiro et al 2012, Desiro et al 2014, Naumann et al 2010). The genomic sequences have confirmed that none of the MRE populations harbours genes involved in the biosynthesis of a Gram-positive cell wall. The nature of the unusual wall-like substance seen in the micrographs remains to be determined, and may possibly be a product of the AMF host.

Although MRE have a reduced genome, and seemingly limited metabolic capabilities, more than half of their genomes encode orphan genes (Table 3.1), with no known functional orthologous groups. Thus, MRE may harbour novel functional pathways that have yet to be characterized.

Host-interactive MRE proteins

The MRE genomes from all three hosts encode proteins with domains that are predominantly used for protein-protein interactions in eukaryotes, specifically ankyrin repeats, ANK, and leucine rich repeats, LRR (Table 3.2). In all these cases, the functions of the genes are unknown; annotations based on amino acid sequence similarity to proteins in the NCBI databases (recovered by BLASTp) revealed similarities to either hypothetical or orphan proteins. ANK proteins in bacteria are generally found in endosymbionts, where they interact with host proteins to have a variety of functions, including differential host transcription, host cytoskeletal changes, and degradation of host proteins (Al-Khodor et al 2010, Pan et al 2008, Richards et al 2013, Schmitz-Esser et al 2010). LRR proteins in bacteria are also relatively rare, but have been found in pathogens of eukaryotic hosts. LRR in bacteria are known to interact with host proteins, which have a variety of effects including adhesion, internalization, and virulence, though in most cases, the actual effect is yet unknown (Bierne et al 2007, Brinster et al 2007, Ishida et al 2014).

Table 3.2. MRE proteins with eukaryotic domains.

	MRE-RC		MRE-RV		MRE-CE	
	Gene loci	**Pfam ID**	**Gene loci**	**Pfam ID**	**Gene loci**	**Pfam ID**
ANK	-	-	contig 7 orf 38	PF12796.2	contig 21 orf 3	PF12796.2
	-	-	contig 7 orf 50	PF12796.2	-	-
	-	-	contig 33 orf 1[c]	PF12796.2	-	-
	-	-	contig 33 orf 6	PF12796.2	-	-
	-	-	contig 73 orf 3	PF12796.2	-	-
	-	-	contig 82 orf 9[c]	PF12796.2	-	-
LRR	contig 1 orf 38[a2]	PF13855.1	contig 1 orf 58	PF12799.2	contig 1 orf 63[f]	PF00560.28
	contig 1 orf 184	PF13855.1	contig 1 orf 97[d]	PF12799.2	contig 1 orf 67[g]	PF13504.1
	contig 2 orf 155	PF00560.28	contig 1 orf 172	PF12799.2	contig 1 orf 100[g]	PF13504.1
	contig 6 orf 48[a]	PF13855.1	contig 2 orf 32	PF12799.2	contig 1 orf 104[f]	PF00560.28
	contig 7 orf 63[b]	PF12799.2	contig 3 orf 80	PF12799.2	contig 12 orf 15	PF12799.2
	contig 8 orf 62[b]	PF12799.2	contig 7 orf 38	PF12796.2	contig 21 orf 3	PF00560.28
	contig 10 orf 23[a]	PF13855.1	contig 7 orf 50	PF12796.2	contig 23 orf 7	PF12799.2
	contig 11 orf 30	PF12799.2	contig 10 orf 14[d]	PF12799.2	contig 25 orf 27	PF12799.2
	-	-	contig 10 orf 48	PF13504.1	-	-
	-	-	contig 16 orf 1	PF12799.2	-	-
	-	-	contig 18 orf 15	PF12799.2	-	-
	-	-	contig 20 orf 31	PF13855.1	-	-
	-	-	contig 31 orf 15	PF12799.2	-	-
	-	-	contig 33 orf 1[e]	PF12796.2	-	-
	-	-	contig 33 orf 6	PF12796.2	-	-
	-	-	contig 44 orf 3	PF12799.2	-	-
	-	-	contig 44 orf 4	PF12799.2	-	-
	-	-	contig 58 orf 10	PF12799.2	-	-
	-	-	contig 73 orf 3	PF12796.2	-	-
	-	-	contig 74 orf 5	PF12799.2	-	-
	-	-	contig 82 orf 9[e]	PF12796.2	-	-
	-	-	contig 85 orf 3	PF13855.1	-	-

Superscripts with the same letter designations indicate duplicates. [a2] slight DNA modifications at the 3' end.

MRE genes of fungal origin

The MRE genomes appear to contain many genes that have been horizontally transferred from their AMF host, representing 5% of total CDS in the MRE-RC metagenome, 3% in MRE-RV, and 4% in MRE-CE (Supplementary Table 3.3). These genes were identified based on amino acid sequence similarity to proteins in the NCBI databases (recovered by BLASTp), with the majority of high similarity genes found in the AMF species, *Rhizophagus irregularis*. *R. irregularis* is the only AMF species with a sequenced genome (Tisserant et al 2013), and is also one of the few AMF that do not harbour MRE (Naumann et al 2010). Unfortunately, none of the genomes of the AMF host species used in this study have yet been sequenced. The potential genes of fungal origin were confirmed through PCR to be true components of the MRE metagenomic contigs. Though most of these products are described only as "hypothetical" and their functions are unknown, the vast majority contained domains involved in protein-protein interactions (Pfam domains LRR & ANK) or signal transduction as tyrosine protein kinases (Pfam domain Pkinase_Tyr). Interestingly, close to half of the ANK & LRR proteins described previously, are derived from AMF (Figure 3.3). Our phylogeny reconstructions revealed that the various LRR proteins in all three MRE populations cluster within the same clade as LRR proteins of AMF origin (*R. irregularis*).

In all three MRE populations, hypothetical proteins with the AIG1 domain are also common. The AIG1 domain was initially described in Arabidopsis, where AIG1 and AIG2 proteins were shown to be involved in plant resistance to bacteria (Reuber and Ausubel 1996). The AIG1 domain is now associated with proteins of the family IAN (immune-associated nucleotide-binding protein), or GIMAP (GTPase of immunity-associated proteins), functionally uncharacterized GTP-binding proteins involved in plant development, plant responses to biotic

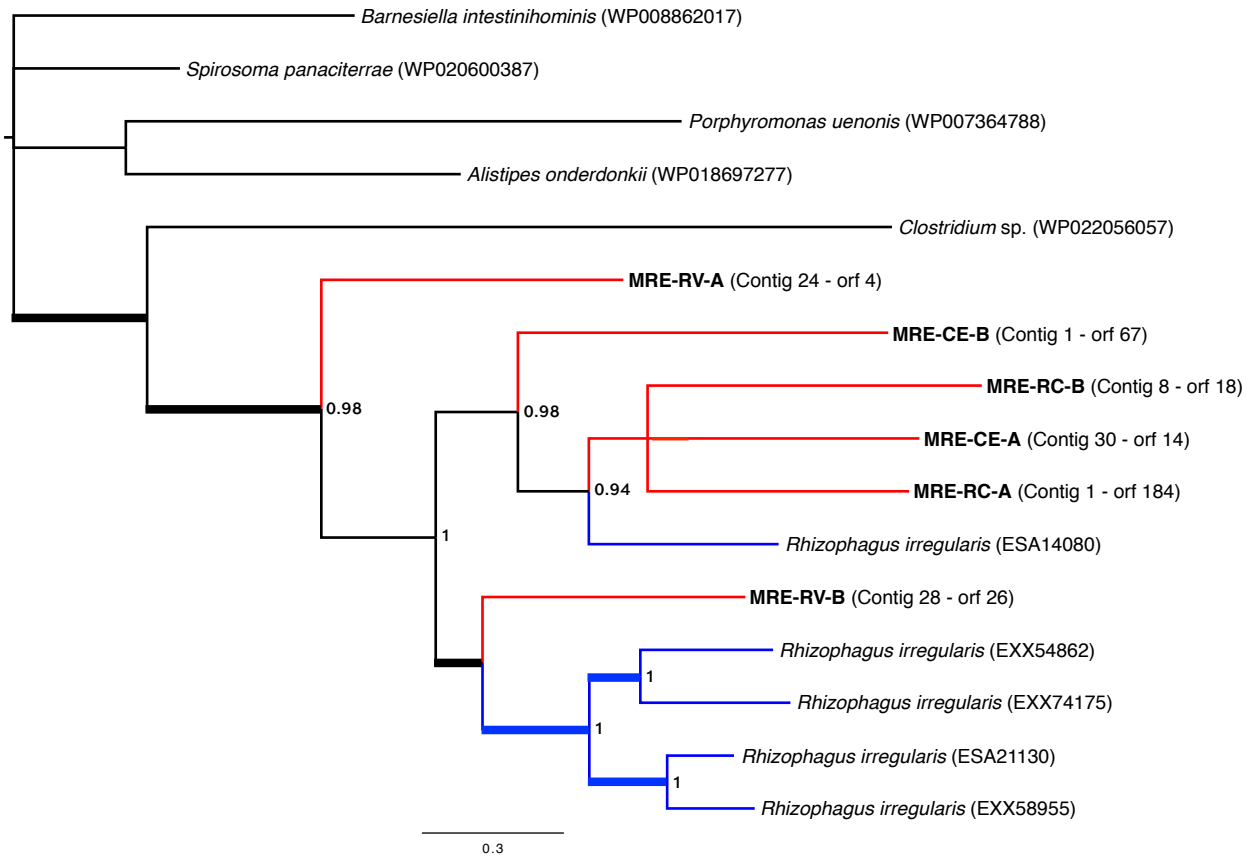


Figure 3.3. Phylogenetic relationship of LRR proteins in MRE. Two LRR proteins from each of the three MRE populations (bolded) were selected for phylogenetic analysis with their closest BLASTp hits. Posterior probabilities greater that 0.90 are shown at nodes. Branches with Maximum Likelihood bootstrap support over 70% are thickened. Branches leading to LRR sequences from MRE are coloured in red and branches leading to LRR sequences from AMF are coloured in blue.

and abiotic stresses, as well as in immunity of vertebrates (Wang and Li 2009). The role of these GTPase proteins in the MRE is unknown.

In the MRE-RV population only, there are three proteins containing the HET (heterokaryon incompatibility) domain. This domain is conserved in Ascomycota fungi, and is involved in the programmed cell death of fungal cells when genetically incompatible hyphae fuse, a phenomenon known as vegetative incompatibility (Paoletti and Clave 2007). However, vegetative incompatibility involving HET proteins has not been observed in AMF and the role of these proteins in MRE remains unclear.

Finally, in both MRE-RV and MRE-CE, sentrin/SUMO proteases have been acquired from the AMF hosts. SUMO proteins are members of the ubiquitin-like family, and are covalently conjugated to other proteins to affect their stability, localization, or interaction partners, a process termed SUMOylation (Everett et al 2013). The SUMO proteases function by deconjugating the SUMO proteins from their substrates. SUMOylation is a strictly eukaryotic process, not found in prokaryotes. Thus, the SUMO proteases encoded in the MRE genomes must be used to alter the SUMOylation state of their AMF hosts. To confirm, when the SUMOfi (SUMO motif finder) tool was used to scan the MRE proteins for possible SUMOylation targets, none was detected (data not shown). Comparison of one of the sequences of the MRE SUMO proteases with its homologous AMF SUMO protease (protein accession ESA14994/nucleotide accession KI282387), revealed that the transfer originally included the full gene sequence with introns and exons, followed by selective degradation to maintain functionality (Figure 3.4, Supplementary Figure 3.1). Furthermore, the presence of SUMO proteases in both MRE-RV and MRE-CE populations, together with the number of mutations differentiating the MRE

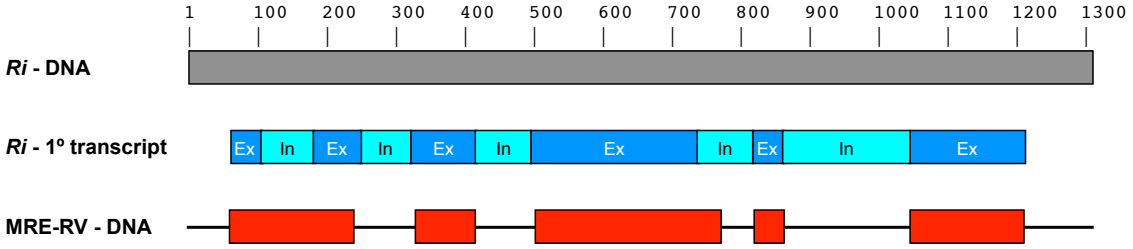


Figure 3.4. Horizontal gene transfer of AMF SUMO protease gene to MRE. The gene encoding the AMF SUMO protease (GenBank Accessions ESA14994/KI282387) is present in the MRE-RV and MRE-CE metagenomes. The gene from MRE-RV, contig 24, orf 13 is shown in this figure as a representation of the overall horizontal gene transfer and subsequent evolution of the gene. Note that the AMF SUMO protease sequence used is that of the species *R. irregularis* (*Ri*), as it is the only AMF sequence available; this sequence may not be the original sequence that was transferred. The full DNA sequence of the SUMO protease gene of AMF (*Ri*) is shown as a grey bar. The corresponding primary transcript sequence is shown below, with exons (Ex) shown as blue bars, and introns (In) shown as pale blue bars. The DNA sequence of the gene from MRE-RV, contig 24, orf 13, is displayed as red bars, with the horizontal line indicating loss of sequences/deletions in the MRE genome. The MRE gene still maintains traces of the introns.

sequence from the AMF sequence, point to the acquisition of the proteases far back during the evolution of the MRE.

Differences between MRE populations

Despite the common ancestry of MRE (Figure 3.1) and their specific niche within the cytoplasm of Glomeromycota, the three MRE populations display differences in genes gained and lost throughout independent evolution. Though we lack fully assembled genomes in all three MRE populations, the sequence coverage of our data is well over 1000x, assuming an overestimated genome size of 1 Mb. Given the estimate that a single species genome can be accurately assembled from a complex metagenome with at least a 20x coverage (Luo et al 2012), we believe that the patterns of gene presence/absence between genomic assemblies of individual MRE populations is reflective of a true absence of those genes, rather than of the insufficient sequence coverage or their intractable nature.

Though the hosts of MRE are all members of Glomeromycota, we expect that divergent MRE populations experience selective pressures specific to their AMF hosts and evolve differently. As apparent in the patterns of presence/absence of genes discussed in previous sections (plectroviral genes and SUMO protease in MRE-RV and MRE-CE populations, and HET domain proteins in the MRE-RV population), each gene's maintenance in the population will depend on its importance for MRE interaction with the specific AMF host species. Other notable differences between MRE populations include the presence of CRISPR-associated Cas1/2 proteins in the MRE-RV and MRE-RC population, but not in the MRE-CE population. The CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins) systems are bacterial defense systems against foreign DNA, such as bacteriophages,

plasmids, and non-self foreign chromosomal DNA (Gasiunas et al 2014, Sampson and Weiss 2013). Furthermore, the CRISPR-Cas system may also be involved in DNA repair and genome evolution (Westra et al 2014). The loss of the CRISPR system in MRE-CE may play a role in the increased genetic diversity seen in this population, similar to increased evolution in the bird pathogen, *Mycoplasma gallisepticum,* associated with the loss of the CRISPR-Cas system (Delaney et al 2012).

Taxonomic position of MRE

Reconstructions of the mollicute phylogenetic history based on 16S rRNA gene sequences are known to be problematic, compared to other groups of bacteria (Thompson et al 2011). Therefore, to resolve MRE relationship with other lineages of the class Mollicutes, we performed a phylogenetic reconstruction using concatenated amino acid sequences of 19 genes (Figure 3.5), based on the genes used in the Genomic Encyclopaedia of Bacteria and Archaea project (Wu et al 2009). We excluded from the analysis the MRE-CE population, as its genetic diversity is likely to confound the results. The multi-gene phylogeny clustered MRE with the members of the family Mycoplasmataceae, suggesting that MRE share a closer relationship with the *Mycoplasma* species than suggested by the 16S rRNA phylogeny (Figure 3.1).

3.5 - DISCUSSION

Our study provides fundamental insights into the biology of a group of endobacteria widely associated with Glomeromycota, the most common and oldest symbionts of plants (Smith and Read 2008). Because of the role of AMF in mineral nutrient and carbon cycles, this tripartite association is one of the most important symbioses on the planet. While the molecular evolution

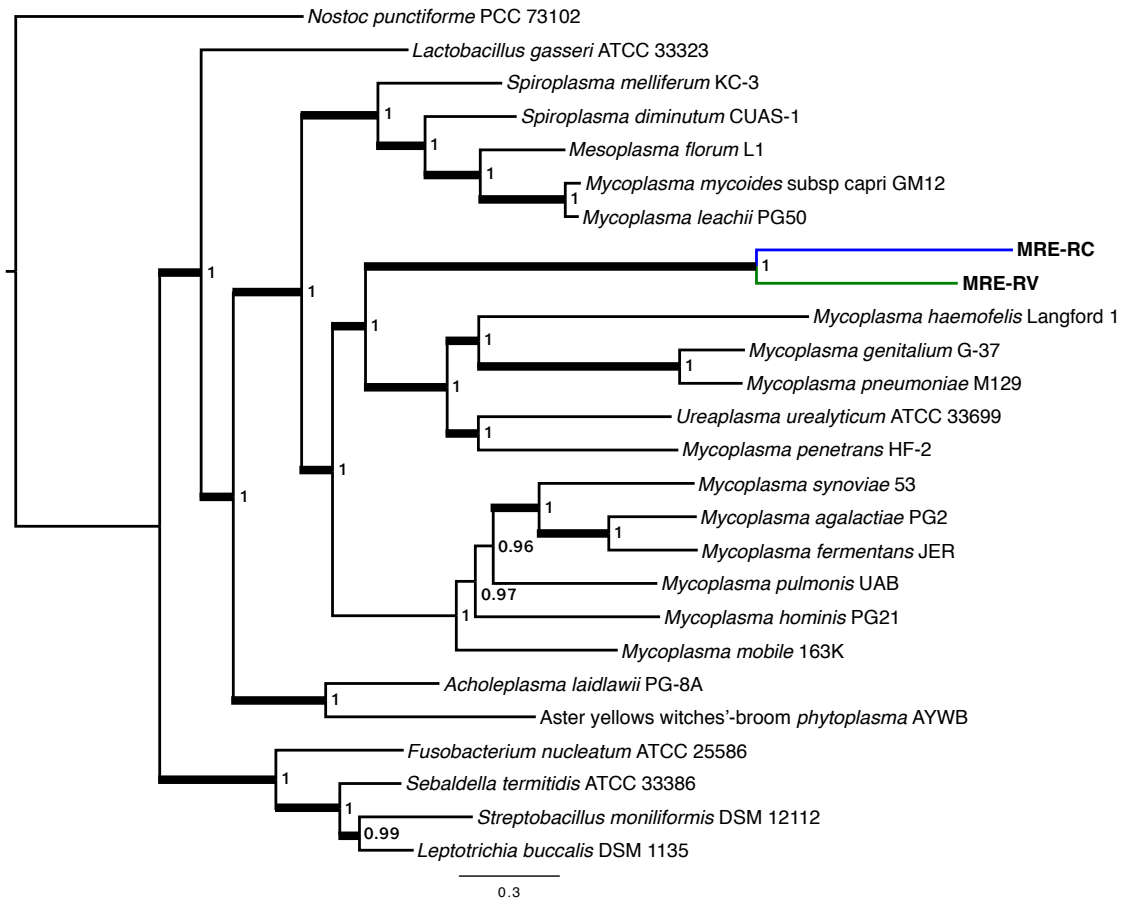


Figure 3.5. Multi-gene phylogeny reveals shared ancestry of MRE and the members of the Mycoplasmataceae family. Bayesian reconstruction was performed on the concatenated amino acid sequences of the following genes: *dnaG, infC, nusA, rplA, rplB, rplC, rplE, rplF, rplM, rplN, rplP, rplT, rpmA, rpsB, rpsC, rpsE, rpsJ, rpsS,* and *smpB.* Posterior probabilities greater that 0.90 are shown at nodes. Branches with Maximum Likelihood bootstrap support over 70% are thickened.

patterns suggest that MRE may be parasites of Glomeromycota (Toomer et al 2014), our present genomic data indicate that MRE are metabolically dependent on AMF and the association between the partners is highly coevolved and evolutionarily stable. Moreover, MRE genomic features imply rampant inter-domain horizontal gene transfer and selection for genes involved in creating genomic diversity.

MRE genome evolution

Genome reduction is one of the hallmarks of endosymbiont reproductive dependence on the host. In heritable mutualists of eukaryotic hosts, genome reduction is a degenerative process related to the small effective size of endosymbiont populations, resulting from population bottlenecks that punctuate transmission between host generations, population subdivision, and clonality (Moran et al 2008). In populations of small effective size, genetic drift is magnified relative to natural selection, which leads to accumulation of slightly deleterious mutations, followed by genome erosion, and contraction. Only the endosymbiont genes important to the provision of goods and services essential to the host are retained owing to the host-level selection (Canbäck et al 2004). The minimal genomes of vertically transmitted MRE present a stark contradiction to this expectation. While they accumulate mutations at a fast pace (Toomer et al 2014), their evolution seems to be driven by mechanisms that are distinctly different from those observed in heritable mutualists. The phylogenetic position of MRE suggests that they are derived from the animal-associated mycoplasmas in the Mycoplasmataceae family. Consistent with this origin, MRE genomes retain the ancestral Mycoplasmataceae propensity for mutation accumulation and gene loss as well as the features responsible for genome plasticity. Rapid mutation accumulation in mycoplasmas is related to the loss of the DNA polymerase proofreading ability conferred by the

polymerase subunits ε (*dnaQ/mutD*) and θ (*holE*) (Kelman and O'Donnell 1995), which are also missing from the MRE genomes. The features responsible for genome plasticity include recombination genes (Moran et al 2008) as well as mobile elements that facilitate recombination (Bischof et al 2006, Li et al 2011, Szczepanek et al 2010). Consistent with the existence of these genomic features in the MRE genomes is evidence of recombination detected in MRE populations using population genetics tools ((Desiro et al 2014, Toomer et al 2014) and unpublished data). The role of recombination in restoration of high fitness genotypes and facilitation of adaptation is well established (Allen et al 2009, Hartfield et al 2012, Muller 1964). It is, therefore, clear that the minimal genomes of MRE are not exclusively products of degenerative evolution. Instead, their degradation due to rapid mutation accumulation is countervailed by the mechanisms of recombination that restore high fitness genotypes and contribute to adaptation to different host cellular environments. Within this context, the genes involved in biosynthesis of metabolites that are readily available in the host environment are expected to experience reduced selective pressure to maintain functionality, and accumulate mutations, leading to gene nonfunctionalization, and loss. What is more, their loss will, in fact, confer fitness advantage through the conservation of cellular energy. MRE genome evolution seems to conform to this adaptive gene loss scenario. A similar scenario, named the Black Queen Hypothesis (BQH), has been recently proposed to describe genome contraction in free-living pelagic bacteria that benefit from costly metabolites released to the environment by other members of the microbial community (Morris et al 2012). One key difference between adaptive genome contraction in MRE (and other mycoplasmas) versus free-living bacteria is the source of costly metabolites; *i.e.* the host in the case of mycoplasmas versus other microbes in the case of free-living bacteria. The other, a more subtle difference, involves the rate at which mutations

accumulate in the genome. Mycoplasmas have one of the highest mutation rates among bacteria owing to the loss DNA repair mechanisms. For example, in *Mycoplasma mycoides* subsp. *mycoides,* mutation rate was estimated at 5×10^{-7} substitutions per site per year (Dupuy et al 2012). In contrast, in free-living pelagic *Prochlorococcus,* mutation rate does not seem to differ from the rates observed in other free-living bacteria, such as *Escherichia coli* (Osburne et al 2011), with mutation rate of 4.5×10^{-9} substitutions per site per year (Ochman et al 1999). Such disparity in mutation rates suggests that superfluous genes can be eliminated from the MRE genomes at a very fast pace whereas retention of essential genes necessitates high rates of recombination to maintain their functionality in the population. This interplay between mutation accumulation and recombination is likely the underlying cause behind the conspicuous genome plasticity apparent in MRE.

MRE-Glomeromycota coevolution

The gene content of the highly reduced MRE genomes suggests that these endobacteria are metabolically dependent on their AMF hosts. As with other members of the Mollicutes class, the limited biosynthetic capabilities of the MRE minimal genomes indicates an intimate relationship with the eukaryotic host. However, unlike most other Mollicutes, MRE reside in the intracellular compartment of their hosts and are transmitted vertically, which likely creates distinct selective pressures.

As a consequence of coevolution, MRE genomes contain a substantial proportion of genes likely involved in interacting with AMF host regulatory networks. Furthermore, many of these genes were acquired horizontally from AMF. Horizontal gene transfer (HGT) in bacteria is a relatively common and important factor in prokaryotic evolution; it is estimated that anywhere

from 1 to 15% of genes in a bacterial genome are due to horizontal transfer (Garcia-Vallve et al 2000). Inter-domain HGT events from eukaryotes to bacteria, such as the ones evidenced in the MRE-AMF system, though not as common, are well documented between organisms that live in close association. For instance, *Legionella* possess many genes acquired via HGT from aquatic protozoa, and these genes are known to allow the bacteria to replicate within eukaryotic hosts (Gomez-Valero et al 2013). Expansin genes, found in certain bacterial species, are another example of inter-domain HGT; these genes were acquired from plants to aid bacteria mediate microbial-plant interactions (Nikolaidis et al 2014). Animal α-amylases are found in many unrelated bacterial species, a product of multiple independent HGT events from various animal donors (Da Lage et al 2004). Clearly, HGT events between closely associated bacteria and their hosts, as in the case of MRE and Glomeromycota, are important coevolutionary mechanisms that allow the recipient organisms to form more intimate associations with their hosts.

One notable example of host manipulation by MRE involves the HGT of SUMO proteases by the bacteria to alter SUMOylation levels of their AMF hosts, a strategy most likely used to change host cytoplasmic conditions in order to benefit endobacterial fitness. The manipulation of host SUMOylation system by bacteria is rare, but has been found in select pathogens. For example, the XopD protein of *Xanthomonas campestris*, a pathogen of tomato and pepper plants, is delivered into the plant's cytosol where it translocates to the nucleus and deSUMOylates host nuclear proteins (Hotson et al 2003). The direct effect of XopD deSUMOylation is unknown, but is presumed to aid in the pathogen's virulence. *Listeria monocytogenes*, a human pathogen capable of intracellular invasion, was shown to decrease host SUMOylation levels by degrading host Ubc9 proteins, an essential enzyme of the SUMOylation machinery (Ribet et al 2010). The decreased SUMOylation levels of host cells lead to increased

bacterial invasion and replication. Similar to SUMOylation, neddylation is a process of the ubiquitin-like family, and involves the conjugation of NEDD8 to substrate proteins to modify their functions. The process of deconjugating NEDD8 from their substrates is known as deneddylation (Rabut and Peter 2008). Enteropathogenic *E. coli* has evolved a method to increase deneddylation of their host cells, leading to their cell cycle arrest and reduction of epithelial turnover; this allows for increased bacterial colonization of the intestinal mucosa (Ashida et al 2014, Jubelin et al 2010). Modifications of host ubiquitin systems by bacterial pathogens has been well documented, and this exploitation has been shown to maximize bacterial proliferation through a variety of methods, including inhibition of host defense and nutrient acquisition (Ashida et al 2014). It is clear that ubiquitin and ubiquitin-like systems of eukaryotic hosts are important targets for many bacterial species. In all cases studied, the hijacking of these system leads directly to increased bacterial proliferation. The acquisition of SUMO proteases by MRE is likely an important strategy allowing for their intracellular proliferation. However, the apparent lack of SUMO proteases in the MRE-RC population indicates that the hijacking of AMF SUMOylation is beneficial, but not essential to MRE survival in the AMF cytoplasm.

It is clear that HGT is an important factor in the coevolution of host-parasite or host-mutualist interactions. The close proximity of the partners creates a hotbed for genetic exchanges, and the genes that favour the interaction are maintained. Even microbial communities that simply reside in the same niche are known to undergo HGT, which inevitably drives their evolution and niche-specific fitness (Juhasz et al 2014, Xiao et al 2011, Zhang et al 2014). The residence of MRE within the cytoplasm of AMF provides them with a constant pool of fungal genes available for HGT, and partner coevolution is evident in the numerous fungal

genes in the MRE genome. This inter-domain HGT has likely allowed for the persistence of MRE in the various AMF groups.

The nature and evolutionary stability of the MRE-Glomeromycota association

The phylogenetic position of MRE as a sister lineage of the *Mycoplasma pneumoniae* group in the family Mycoplasmataceae, together with the Mycoplasmataceae origin at 410 million mya (Maniloff 2002) suggest a considerable antiquity of the MRE-Glomeromycota association. All the same, Glomeromycota were likely free of MRE during their early evolutionary history, which extends to at least 460 mya, as evidenced by fossil record (Redecker et al 2000). The great age of the MRE-Glomeromycota association and its extant ubiquity suggest that this symbiosis is evolutionarily and ecologically stable despite the complete metabolic dependence of MRE on the AMF hosts and the ability of some AMF fungal species to exist without MRE infection.

MRE share with other *Mycoplasma* species a collection of genomic features, including metabolic dependence on the host as well as mechanisms ensuring genomic diversity (Marenda 2014), including recombination machinery and transposable elements. However, unlike the present day *Mycoplasma* species that rely primarily on horizontal dispersal, MRE are for the most part transmitted vertically from one host generation to the next, with occasional instances of horizontal dispersal (Toomer et al 2014). The shift from predominantly horizontal to predominantly vertical transmission could be expected to eliminate selective pressures that maintain the mechanisms underlying genomic diversity. Such losses of recombination machinery are common in heritable mutualists of insects that are essential to the survival of their hosts (Moran et al 2008). These essential endosymbionts evolve under genetic drift dominated

constrains of small effective population sizes, resulting from mutualistic lifestyle and vertical transmission. As a consequence, evolution of heritable mutualists is driven entirely by the selective pressures experienced by their hosts (Canbäck et al 2004, O'Fallon 2008). The fact that MRE retain the mechanisms responsible for genomic diversity suggests that, despite MRE predominantly vertical transmission and in contrast to essential mutualists, these mechanisms are selected for. These selective pressures are most likely generated by the host defensive responses to the inevitable exploitation by MRE and/or by the demands of competition against other MRE genotypes for the host resources. Both of these processes can be expected to drive the coevolutionary race between MRE and their Glomeromycota hosts and fuel the demand for continuous adaptation. Consequently, longevity of the MRE-Glomeromycota association appears to depend on the mechanisms that generate MRE genomic diversity.

No functional data are available on the nature of the association between MRE and their Glomeromycota hosts and no direct clues have emerged from the genomic data that we generated in the present study. However, the retention of the genetic mechanisms that underlie genomic diversity in MRE suggests that MRE may be antagonists of their fungal hosts. This suggestion is consistent with the patterns of molecular evolution detected in the study of MRE population structure (Toomer et al 2014). It is also supported by observations from the *Wolbachia* reproductive parasites of insects. Like MRE, *Wolbachia* engage in genetic recombination and their genomes harbour high numbers of mobile elements (Brelsfoard et al 2014, Leclercq et al 2011, Sloan and Moran 2013).

Trajectories of MRE evolution

The study of the MRE populations associated with three distinct hosts allowed us to compare their evolutionary trajectories. For instance, the acquisition of the HET domain genes was only seen in MRE-RV, while the SUMO protease genes were seen only in the MRE-RV and MRE-CE population. While the role of the HET domain proteins is unknown, the SUMO proteases seem to be highly important for MRE, especially since the gene originates from the AMF host genome. Given the apparent significance of SUMO proteases, it is intriguing that the MRE-RC population does not seem to harbour them. The AMF hosts CE and RC are more closely related than CE and RV, which points to either the acquisition of SUMO proteases by MRE prior to the diversification of the three AMF host species, followed by the loss of this gene in the MRE-RC lineage, or the acquisition of the genes by MRE-RV and MRE-CE independently. Clearly, more MRE populations must be studied for the presence/absence of these genes to truly understand the evolution of the MRE genes of fungal origin.

Experimental challenges

The MRE-Glomeromycota symbiosis poses considerable experimental challenges, including uncultivability of MRE (Chapter 2) and limited cultivability of AMF (Cranenbrouck et al 2005). However, our MRE genomic data together with the recently published genome of *R. irregularis* (Tisserant et al 2013) create foundations for exploring the biology of these organisms through the host-induced-gene-silencing in AMF (Helber et al 2011) and *ex vivo* functional studies of MRE genes of interest (Ishii et al 2013). The organisms involved in this symbiosis are important components of the global mineral nutrient and carbon cycles. They also constitute a unique model for the study of endosymbiotic systems and endobacterial evolution.

3.6 - CONCLUSION

This is the first study to explore the biological capabilities and the mechanisms of co-evolution of MRE with Glomeromycota. Though MRE existence was known for close to half a century (Mosse 1970), the ubiquity of the MRE-Glomeromycota association and its biological importance are only being discovered. Based on our metagenomic study, the MRE-Glomeromycota relationship is highly coevolved and of great, potentially antagonistic, significance to the partners. As microbiome studies revealed the importance of microbial communities in higher eukaryotes, our work uncovered a tightly knit network of interactions between fungi and their own microbial communities. We thus created the foundation to study the obscure yet key member of the Glomeromycota-associated microbial community.

3.7 - MATERIALS & METHODS

AMF materials. AMF species *Claroideoglomus etunicatum* CA-OT135 spores were originally obtained from maize experimental fields known as Oxford Tract in California (den Bakker et al 2010). The spores were maintained *in vitro*, grown on root-inducing T-DNA-transformed carrot roots (clone DC2). Roots and AMF were grown on minimal medium modified with 10 mM 2-(4-morpholine)-ethanesulfonic acid (pH 6.0; Sigma-Aldrich) and solidified with Phytagel™ (Sigma-Aldrich) (den Bakker et al 2010), and incubated at 28ºC. *Rhizophagus clarus* NB112A was obtained from the International Culture Collection of Arbuscular Mycorrhizal Fungi (INVAM), and maintained *in vitro*, on root-inducing T-DNA-transformed chicory roots, grown on MSR medium solidified with Phytagel™ (Cranenbrouck et al 2005), and incubated at 28ºC. *Racocetra verrucosa* VA103A was also obtained from INVAM and its spores were used directly after extraction from the potting medium and surface decontamination.

MRE sample preparation and Illumina sequencing. *C. etunicatum* CA-OT135 and *R. clarus* NB112A spores were extracted from *in vitro* cultures by dissolving the Phytagel™ medium in 10 mM sodium citrate buffer (pH 6.0; Fisher Scientific) at 30ºC for 20 min. *R. verrucosa* VA103A spores were extracted from potting medium by wet sieving and sucrose centrifugation, as described previously (den Bakker et al 2010). The extracted spores were surface sterilized using hydrogen peroxide and chloramine-T (Mondo et al 2012), followed by 24 h of incubation in antibiotic solution (2% streptomycin sulphate and 1% gentamicin sulphate). The MRE populations were extracted from their respective AMF hosts by manually crushing 100 spores of each AMF species in sterile dH_2O, and filtered through a 1 µm filter. For MRE-RV and MRE-CE, the filtered samples were each divided into 5 parts, and whole genome amplification using Illustra™ GenomiPhi V2 Amplification Kit (GE Healthcare) was performed following the manufacturer's protocol, and amplified products were pooled again post-amplification. For MRE-RC, the genomic DNA was extracted directly from the filtrate using QIAamp DNA Micro Kit (Qiagen) using the protocol for Isolation of Genomic DNA from Laser-Microdissected Tissues. The TruSeq DNA library preparation and 2 × 100 paired end HiSeq sequencing was performed by the Cornell University Life Sciences Core Laboratories Center.

Assembly of raw reads. The raw reads were analyzed for quality and potential problems during sequencing using the freely available FASTQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Based on the analysis, FASTX-TOOLKIT (http://hannonlab.cshl.edu/fastx_toolkit/index.html) was used to trim the first few nucleotides of all reads and remove reads that contained an indistinguishable (N) nucleotide.

Furthermore, FASTQ-MCF (https://code.google.com/p/ea-utils/) was used to remove nucleotides with poor quality and to remove reads with a CASAVA "Y" flag indicating a failure to pass a quality filter. The resulting reads were interlaced according to their pairs using the FASTQ Interlacer on GALAXY (Goecks et al 2010), in preparation for assembly. The cleaned reads were assembled using MetaVelvet, an assembly program designed for a metagenomic sequences (Namiki et al 2012). Various kmer lengths were tested and used for optimal assembly results. The contigs from MetaVelvet assemblies were further assembled into supercontigs using the Geneious v6 (http://www.geneious.com), using the default settings. Assembled supercontigs were inspected manually for misassemblies.

Annotation of contigs. Coding regions were identified using GLIMMER v.3.02 using the Mycoplasma genetic code (Delcher et al 2007). Annotation of the identified genes was performed using BLASTp (E value cutoff 10^{-4}) against a non-redundant protein database (Altschul et al 1990). Conserved domains were identified using the Pfam database (Finn et al 2014). Functional annotation was assigned using the KEGG Automatic Annotation Server (KAAS) using the bi-directional best hit method to assign KEGG Orthology (KO) identities (Moriya et al 2007). The "Genes Data Set" used in KAAS was manually selected for the MRE: "sce, eco, hin, pae, nme, hpy, rpr, mlo, bsu, mge, ncr, buc, brh, cac, mpu, mpe, mmy, mmo, msy, mcp, mat, uur, ayw, acl, mfl, uma, bja, lsa, mhj, mco, mcd, mho, maa, mpn, uue, pal, bfa, ath, aor, tml" was used. tRNAs were detected using tRNAscan-SE 1.21 (Lowe and Eddy 1997).

SUMOylation targets. The online tool SUMOFI (SUMO Motif Finder) (http://cbg.garvan.unsw.edu.au/sumofi/home.do) was used to scan for SUMOylation targets in the protein sequences of all three MRE populations. The default motif ψKxE was used.

Phylogenetic reconstruction of 16S rRNA sequences. Individual AMF spores were manually crushed in sterile dH_2O, and used as templates for PCR reaction with MRE-specific 16S rRNA primers 109F1 (5'-ACGGGTGAGTAATRCTTATCT -3'), 109F2 (5'-ACGAGTGAGTAATGCTTATCT -3'), 1184R1 (5'- GACGACCAGACGTCATCCTY -3'), 1184R2 (5'- GACGACCAAACTTGATCCTC -3'), and 1184R3 (5'-GATGATCAGACGTCATCCTC -3') (Naumann et al 2010). Phusion® Hot Start Flex DNA Polymerase (New England Biolabs) was used following the manufacturer's directions, with an initial denaturation time of 3 min, a melting temperature of 60ºC, and an extension time of 1 min. The cycle was repeated 20 times. The primers were added to a final concentration of 0.75 µM for 109F1 and 1184R1, and 0.375 µM for 109F2, 1184R2, and 1184R3. The amplicons were purified using QIAquick PCR purification kit (Qiagen), cloned into CloneJET pJET1.2 vector (Thermo Scientific), and transformed into One Shot® Top10 Chemically Competent *E. coli* cells (Invitrogen). Individual colonies were selected and sequenced for their pJET1.2 vector insert. Two representative 16S rRNA sequences from three spores per AMF host species (6 sequences per MRE population) were used for the phylogenetic reconstruction. All non-MRE sequences were obtained from IMG: the Integrated Microbial Genomes database (Markowitz et al 2012). Nucleotide sequences were aligned using MUSCLE (Edgar 2004), and Bayesian phylogenies were constructed using MrBayes v.3.1.2 (Ronquist and Huelsenbeck 2003) with the GTR+I+Γ nucleotide substitution model run for 2,000,000 generations with 25% burn-in.

Phylogenetic reconstruction of LRR and 19 concatenated conserved genes. For LRR phylogeny, the amino acid sequences of two LRR proteins from each MRE population were randomly selected. Non-MRE sequences were obtained from similarity searches based on amino acid sequences against the NCBI database (BLASTp). For the multi-gene phylogeny, amino acid sequences of the following 19 genes were selected and concatenated: *dnaG, infC, nusA, rplA, rplB, rplC, rplE, rplF, rplM, rplN, rplP, rplT, rpmA, rpsB, rpsC, rpsE, rpsJ, rpsS,* and *smpB*. Non-MRE sequences were obtained from IMG (Markowitz et al 2012). For both phylogenies, sequences were aligned with MUSCLE (Edgar 2004), and Bayesian phylogenetic trees were constructed using MrBayes v.3.1.2 (Ronquist and Huelsenbeck 2003) using mixed the I+Γ amino acid substitution model run for 1,000,000 generations with 25% burn-in for the LRR phylogeny, and 2,000,000 generations with 25% burn-in for the multi-gene phylogeny.

Accession Numbers. All contigs generated in this study are deposited in GenBank (pending request for accession numbers at time of dissertation publication).

3.8 - ACKNOWLEDGEMENTS

We thank D. Douds for root-inducing T-DNA-transformed carrot roots (clone DC2). This work was supported by the National Science Foundation grant DEB-0918880 and by the Cornell University Agricultural Experiment Station federal formula funds, Project NYC-153434, received from the National Institutes for Food and Agriculture, U.S. Department of Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the U.S. Department of Agriculture.

3.9 - SUPPORTING INFORMATION

Table S3.1. Comparison of contigs that share regions of identity and gene synteny. The contigs display portions of the chromosome that have undergone chromosomal rearrangement in the MRE populations. The likely cause of rearrangement is indicated.

MRE population	Contigs	Identical ORF range	Length of identity (bp)	Cause of rearrangement
MRE-RC	1 9	1 - 10 60 - 71	4,339	unknown/recombination
	1 10	15 - 38 1 - 23	15,066	unknown/recombination
	2 4 6	25 - 1 128 - 152 75 - 100	10,770	transposase (uncharacterized)
	3 9	15 - 3 73 - 86	4,278	unknown/recombination
	3 12	93 - 120 33 - 66	12,920	unknown/recombination
	4 6	117 - 126 65 - 56	9,027	transposase (IS605)
	6 10 6*	43 - 56 19 - 32 73 - 68	6,610 4,760*	transposase (IS605)
	6 13	1 - 42 1 - 44	19,974	transposase (IS605)
	7 8	18 - 81 18 - 79	35,557	transposase (IS605)
	7 9	91 - 85 2 - 9	2,062	transposase (uncharacterized)
	8 11	1 - 11 10 - 1	5,133	transposases (IS605/IS200)
MRE-RV	1 10	77 - 109 33 - 1	19,757	SSC/recombination
	4 11	37 - 105 1 - 60	23,102	unknown/recombination
	4 23	6 - 31 5 - 28	14,461	SSC/recombination
	5 14	85 - 77 28 - 69	1,700	plectroviral invasion/ high polymorphism
	14 21	50 - 63 5 - 18	4,282	plectroviral invasion?
	17 22	6 - 15 37 - 45	5,691	transposase (IS1595)
	33 82	1 - 4 9 - 6	1,744	transposase (IS605)

MRE-CE	1 1	1 - 78 166 - 89	27,237	inversion
	3 10	7 - 11 47 - 50	2,516	SSC/recombination polymorphism
	3 18	1 - 14 17 - 33	4,675	unknown/recombination
	4 7	49 - 61 38 - 27	7,417	unknown/recombination polymorphism
	4 25	42 - 46 22 - 17	2,340	transposase (uncharacterized)
	6 17	3 - 47 45 - 3	13,046	SSC/recombination
	10 15	11 - 36 1 - 18	9,183	SSC/recombination
	11 31	1 - 21 all; 1 - 13	7,393	polymorphism (divergence)
	18 29	2 - 16 2 - 16	5,541	transposase (uncharacterized)
	35 39	15 - 32 4 - 14	3,831	SSC/recombination

* The length of identity is shorter than the other 2 contigs.

Table S3.2. Genes involved in DNA recombination maintained in the MRE genome.

	MRE-RC	MRE-RV	MRE-CE
recA (homologous recombination)	contig 14 - orf 23	contig 2 - orf 91	contig 57 - orf 1
xerD (XER site-specific recombination)	contig 7 - orf 1	contig 1 - orf 134	contig 1 - orf 15[b]
	contig 9 - orf 17	contig 1 - orf 139	contig 1 - orf 86
	contig 10 - orf 54	contig 4 - orf 74[a]	contig 1 - orf 152[b]
	contig 11 - orf 49	contig 9 - orf 5	contig 2 - orf 79
	-	contig 9 - orf 56	contig 2 - orf 139
	-	contig 10 - orf 35	contig 7 - orf 19
	-	contig 11 - orf 31[a]	contig 33 - orf 27
	-	contig 22 - orf 18	contig 47 - orf 8
	-	contig 35 - orf 26	-
	-	contig 36 - orf 8	-
	-	contig 36 - orf 17	-
	-	contig 49 - orf 10	-
	-	contig 78 - orf 10	-
	-	contig 86 - orf 2	-
xerC (XER site-specific recombination)	-	contig 4 - orf 106	contig 11 - orf 42
	-	contig 27 - orf 6	contig 35 - orf 22[c]
	-	-	contig 39 - orf 9[c]

Superscripts with the same letter designations indicate duplicates.

Table S3.3. MRE genes of fungal origin.

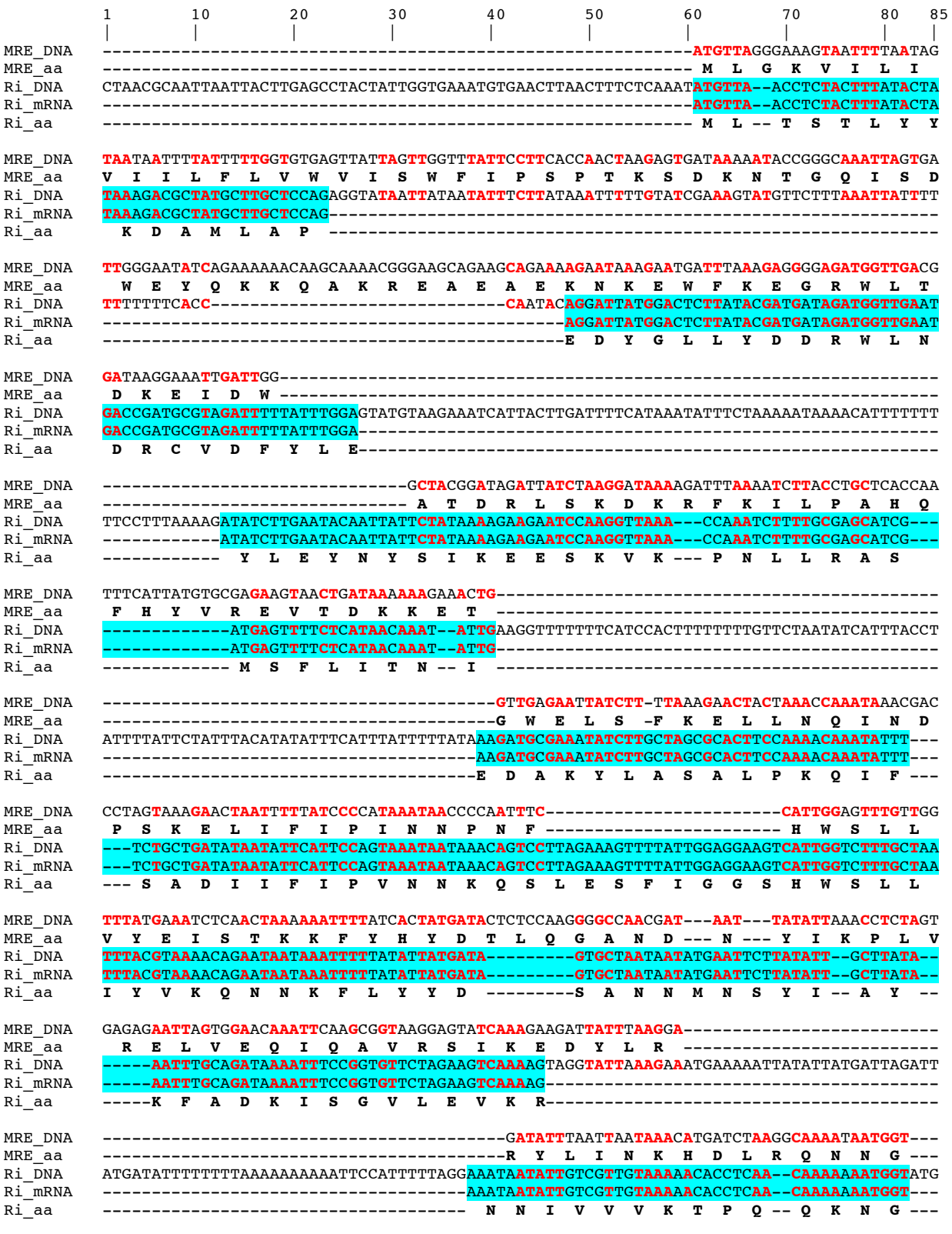
Locus	BLASTp hit	GenBank accession	Hit organism	Pfam domain (MRE gene)
MRE-RC				
Contig 1 - orf 4	hypothetical protein	ERZ99841	*Rhizophagus irregularis* (AMF)	tRNA_U5-meth_tr
Contig 1 - orf 6	hypothetical protein	ERZ99841	*R. irregularis* (AMF)	Pfam-B_18852
Contig 1 - orf 7	hypothetical protein	ERZ99841	*R. irregularis* (AMF)	-
Contig 1 - orf 26	hypothetical protein	ERZ94661	*R. irregularis* (AMF)	DNA_binding_1
Contig 1 - orf 38	hypothetical protein	ESA14080	*R. irregularis* (AMF)	LRR_8
Contig 1 - orf 160	hypothetical protein	ESA20877	*R. irregularis* (AMF)	AIG1
Contig 1 - orf 184	hypothetical protein	ESA20166	*R. irregularis* (AMF)	LRR_1, LRR_8
Contig 1 - orf 192	hypothetical protein	ESA20791	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 1 - orf 231	hypothetical protein	ESA22813	*R. irregularis* (AMF)	Fructosamin_kin
Contig 2 - orf 91	hypothetical protein	ESA23675	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 2 - orf 94	hypothetical protein	ESA00890	*R. irregularis* (AMF)	-
Contig 2 - orf 102	hypothetical protein	ESA06709	*R. irregularis* (AMF)	-
Contig 2 - orf 108	hypothetical protein	ESA09460	*R. irregularis* (AMF)	-
Contig 2 - orf 114	hypothetical protein	ESA23675	*R. irregularis* (AMF)	Pkinase
Contig 2 - orf 120	hypothetical protein	ESA23432	*R. irregularis* (AMF)	-
Contig 2 - orf 155	hypothetical protein	ESA20165	*R. irregularis* (AMF)	LRR_1
Contig 2 - orf 159	hypothetical protein	ESA10103	*R. irregularis* (AMF)	Abhydrolase_2
Contig 3 - orf 23	hypothetical protein	ESA09631	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 3 - orf 24	hypothetical protein	ESA12031	*R. irregularis* (AMF)	-
Contig 3 - orf 39	hypothetical protein	ESA03387	*R. irregularis* (AMF)	LRAT
Contig 4 - orf 13	hypothetical protein	ESA20877	*R. irregularis* (AMF)	AIG1
Contig 4 - orf 21	hypothetical protein	ESA20877	*R. irregularis* (AMF)	AIG1
Contig 4 - orf 40	hypothetical protein	ESA03923	*R. irregularis* (AMF)	LRAT
Contig 5 - orf 58	hypothetical protein	ESA10103	*R. irregularis* (AMF)	Abhydrolase_2
Contig 5 - orf 85	hypothetical protein	ERZ99664	*R. irregularis* (AMF)	-
Contig 5 - orf 87	hypothetical protein	ERZ99664	*R. irregularis* (AMF)	-
Contig 6 - orf 48	hypothetical protein	ESA14080	*R. irregularis* (AMF)	LRR_8
Contig 6 - orf 52	hypothetical protein	ESA11624	*R. irregularis* (AMF)	Phosphodiest
Contig 6 - orf 53	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 6 - orf 68	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 6 - orf 69	hypothetical protein	ESA11624	*R. irregularis* (AMF)	Phosphodiest
Contig 6 - orf 73	hypothetical protein	ESA14080	*R. irregularis* (AMF)	-
Contig 7 - orf 7	hypothetical protein	ESA22250	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 7 - orf 9	hypothetical protein	ESA21130	*R. irregularis* (AMF)	Pfam-B_5380
Contig 7 - orf 18	hypothetical protein	ESA14080	*R. irregularis* (AMF)	-
Contig 7 - orf 83	hypothetical protein	ESA19770	*R. irregularis* (AMF)	-
Contig 8 - orf 18	hypothetical protein	ESA14080	*R. irregularis* (AMF)	-
Contig 9 - orf 11	hypothetical protein	ESA19770	*R. irregularis* (AMF)	-
Contig 9 - orf 64	hypothetical protein	ERZ99841	*R. irregularis* (AMF)	tRNA_U5-meth_tr
Contig 9 - orf 66	hypothetical protein	ERZ99841	*R. irregularis* (AMF)	Pfam-B_18852
Contig 9 - orf 67	hypothetical protein	ERZ99841	*R. irregularis* (AMF)	-
Contig 10 - orf 10	hypothetical protein	ERZ94661	*R. irregularis* (AMF)	DNA_binding_1
Contig 10 - orf 23	hypothetical protein	ESA14080	*R. irregularis* (AMF)	LRR_8
Contig 10 - orf 27	hypothetical protein	ESA11624	*R. irregularis* (AMF)	Phosphodiest
Contig 10 - orf 28	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 10 - orf 35	hypothetical protein	ESA14080	*R. irregularis* (AMF)	-
Contig 10 - orf 45	hypothetical protein	ESA14080	*R. irregularis* (AMF)	Pfam-B_5380
Contig 10 - orf 48	hypothetical protein	ESA22250	*R. irregularis* (AMF)	Pkinase_Tyr

Contig 11 - orf 25	hypothetical protein	ESA10766	*R. irregularis* (AMF)	DUF1768
Contig 11 - orf 42	hypothetical protein	ESA14080	*R. irregularis* (AMF)	-
Contig 12 - orf 3	hypothetical protein	ERZ99536	*R. irregularis* (AMF)	-
Contig 15 - orf 1	hypothetical protein	ESA12358	*R. irregularis* (AMF)	Pfam-B_2045
Contig 15 - orf 4	hypothetical protein	ESA12358	*R. irregularis* (AMF)	-
Contig 16 - orf 4	hypothetical protein	ESA07651	*R. irregularis* (AMF)	-
MRE-RV				
Contig 1 - orf 64	hypothetical protein	ESA14359	*R. irregularis* (AMF)	AIG1, Pkinase
Contig 1 - orf 65	hypothetical protein	ESA16792	*R. irregularis* (AMF)	AIG1
Contig 1 - orf 97	hypothetical protein	ESA08063	*R. irregularis* (AMF)	LRR_4, Pkinase_Tyr
Contig 1 - orf 172	hypothetical protein	ESA21130	*R. irregularis* (AMF)	LRR_4, LRR_1
Contig 2 - orf 31	hypothetical protein	ERZ99038	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 2 - orf 32	hypothetical protein	ESA14080	*R. irregularis* (AMF)	LRR_4
Contig 2 - orf 101	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 3- orf 17	hypothetical protein	ESA09071	*R. irregularis* (AMF)	AIG1
Contig 3 - orf 19	hypothetical protein	ESA09071	*R. irregularis* (AMF)	-
Contig 4 - orf 14	hypothetical protein	ESA10211	*R. irregularis* (AMF)	DUF1768
Contig 7 - orf 4	hypothetical protein	ESA00890	*R. irregularis* (AMF)	-
Contig 7 - orf 60	hypothetical protein	ESA02142	*R. irregularis* (AMF)	AIG1
Contig 7 - orf 64	sentrin-specific protease 8	EFA81807	*Polysphondylium pallidum* (cellular slime mould)	Peptidase_C48
Contig 7 - orf 83	hypothetical protein	ESA23675	*R. irregularis* (AMF)	Pkinase
Contig 10 - orf 14	hypothetical protein	ESA08063	*R. irregularis* (AMF)	LRR_4, Pkinase_Tyr
Contig 10 - orf 46	hypothetical protein	ESA10593	*R. irregularis* (AMF)	-
Contig 10 - orf 48	hypothetical protein	ESA09071	*R. irregularis* (AMF)	LRR_7, LRR_1, AIG1
Contig 12 - orf 26	hypothetical protein	ESA17559	*R. irregularis* (AMF)	AIG1, Peptidase_C48
Contig 12 - orf 30	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 12 - orf 31	heterokaryon incompatibility protein	ENH85205	*Colletotrichum orbiculare* (Ascomycota)	HET
Contig 13 - orf 28	hypothetical protein	ESA09071	*R. irregularis* (AMF)	AIG1
Contig 13 - orf 31	hypothetical protein	ESA23675	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 17 - orf 6	hypothetical protein	ESA03396	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 18 - orf 15	hypothetical protein	ESA13621	*R. irregularis* (AMF)	LRR_4, Pkinase
Contig 19 - orf 18	hypothetical protein	ERZ96796	*R. irregularis* (AMF)	-
Contig 19 - orf 25	hypothetical protein	ESA03725	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 20 - orf 26	HET domain protein	EAW13658	*Aspergillus clavatus* (Ascomycota)	HET
Contig 20 - orf 31	hypothetical protein	ESA21130	*R. irregularis* (AMF)	LRR_8
Contig 20 - orf 43	hypothetical protein	ESA09071	*R. irregularis* (AMF)	AIG1
Contig 20 - orf 45	hypothetical protein	ESA09278	*R. irregularis* (AMF)	AIG1
Contig 20 - orf 49	hypothetical protein	ESA09071	*R. irregularis* (AMF)	AIG1
Contig 20 - orf 51	hypothetical protein	ESA09278	*R. irregularis* (AMF)	AIG1
Contig 20 - orf 53	hypothetical protein	ESA02142	*R. irregularis* (AMF)	AIG1
Contig 21 - orf 30	hypothetical protein	ESA03923	*R. irregularis* (AMF)	LRAT
Contig 21 - orf 32	hypothetical protein	ESA09032	*R. irregularis* (AMF)	-
Contig 22 - orf 37	hypothetical protein	ERZ99038	*R. irregularis* (AMF)	Pkinase
Contig 23 - orf 13	hypothetical protein	ESA10211	*R. irregularis* (AMF)	DUF1768
Contig 24 - orf 4	hypothetical protein	ESA21130	*R. irregularis* (AMF)	Pfam-B_5380
Contig 24 - orf 9	hypothetical protein	ESA21130	*R. irregularis* (AMF)	HET
Contig 24 - orf 13	sentrin-specific	EFA81807	*P. pallidum* (cellular slime	Peptidase_C48

	protease 8		mould)	
Contig 25 - orf 38	hypothetical protein	ESA14080	*R. irregularis* (AMF)	Pfam-B_5380
Contig 27 - orf 29	hypothetical protein	ESA04725	*R. irregularis* (AMF)	-
Contig 28 - orf 26	hypothetical protein	ESA21130	*R. irregularis* (AMF)	Pfam-B_1581
Contig 28 - orf 30	hypothetical protein	EPS44510	*Dactylellina haptotyla* (Ascomycota)	Peptidase_C48
Contig 31 - orf 15	hypothetical protein	ESA14080	*R. irregularis* (AMF)	LRR_4
Contig 33 - orf 1	hypothetical protein	EHK49265	*Trichoderma atroviride* (Ascomycota)	Ank_2
Contig 33 - orf18	hypothetical protein	ESA03396	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 33 - orf 25	hypothetical protein	ESA21121	*R. irregularis* (AMF)	-
Contig 38 - orf 14	hypothetical protein	ESA03396	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 44 - orf 3	hypothetical protein	ESA20165	*R. irregularis* (AMF)	LRR_4
Contig 44 - orf 4	HET domain protein	EAW13658	*A. clavatus* (Ascomycota)	LRR_4, HET
Contig 53 - orf 5	hypothetical protein	ERZ99536	*R. irregularis* (AMF)	-
Contig 56 - orf 4	hypothetical protein	ERZ94661	*R. irregularis* (AMF)	DNA_binding_1
Contig 62 - orf 2	hypothetical protein	ESA02142	*R. irregularis* (AMF)	AIG1
Contig 82 - orf 9	hypothetical protein	EHK49265	*T. atroviride* (Ascomycota)	Ank_2
MRE-CE				
Contig 1 - orf 33	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 1 - orf 36	hypothetical protein	ESA23675	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 1 - orf 59	hypothetical protein	ESA14080	*R. irregularis* (AMF)	-
Contig 1 - orf 63	hypothetical protein	ESA21130	*R. irregularis* (AMF)	LRR_1, LRR_7
Contig 1 - orf 67	hypothetical protein	ESA21130	*R. irregularis* (AMF)	LRR_7
Contig 1 - orf 100	hypothetical protein	ESA21130	*R. irregularis* (AMF)	LRR_7
Contig 1 - orf 104	hypothetical protein	ESA21130	*R. irregularis* (AMF)	LRR_1, LRR_7
Contig 1 - orf 108	hypothetical protein	ESA14080	*R. irregularis* (AMF)	-
Contig 1 - orf 133	hypothetical protein	ESA23675	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 1 - orf 134	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 2 - orf 69	hypothetical protein	ESA03387	*R. irregularis* (AMF)	-
Contig 2 - orf 113	hypothetical protein	ESA23432	*R. irregularis* (AMF)	-
Contig 2 - orf 132	hypothetical protein	ESA04490	*R. irregularis* (AMF)	Pfam-B_18364
Contig 4 - orf 30	hypothetical protein	ESA07883	*R. irregularis* (AMF)	malic
Contig 9 - orf 19	hypothetical protein	ESA03396	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 9 - orf 24	sentrin/sumo-specific protease	EER24556	*Coccidioides posadasii* (Ascomycota)	Peptidase_C48
Contig 10 - orf 43	hypothetical protein	ESA07883	*R. irregularis* (AMF)	malic
Contig 11 - orf 17	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 12 - orf 15	hypothetical protein	ESA00099	*R. irregularis* (AMF)	LRR_4, Pkinase_Tyr
Contig 12 - orf 21	sentrin/sumo-specific protease	EER24556	*Coccidioides posadasii* (Ascomycota)	Peptidase_C48
Contig 13 - orf 28	hypothetical protein	ESA08063	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 13 - orf 53	hypothetical protein	ESA19829	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 15 - orf 27	hypothetical protein	ESA08063	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 18 - orf 16	hypothetical protein	ESA14080	*R. irregularis* (AMF)	Pfam-B_8398
Contig 23 - orf 7	hypothetical protein	ESA24131	*R. irregularis* (AMF)	LRR_4, Pkinase_Tyr
Contig 27 - orf 5	hypothetical protein	ESA21121	*R. irregularis* (AMF)	-
Contig 29 - orf 16	hypothetical protein	ESA14080	*R. irregularis* (AMF)	Pfam-B_8398
Contig 30 - orf 14	hypothetical protein	ESA14080	*R. irregularis* (AMF)	Pfam-B_5380
Contig 31 - orf 12	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 32 - orf 2	hypothetical protein	ESA21130	*R. irregularis* (AMF)	Pfam-B_5380
Contig 32 - orf 6	hypothetical protein	ESA11422	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 32 - orf 10	hypothetical protein	ESA04490	*R. irregularis* (AMF)	-

Contig 32 - orf 15	hypothetical protein	ESA02018	*R. irregularis* (AMF)	-
Contig 38 - orf 14	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 38 - orf 25	hypothetical protein	ESA24131	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 40 - orf 6	hypothetical protein	ESA08063	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 41 - orf 14	hypothetical protein	ESA07651	*R. irregularis* (AMF)	-
Contig 47 - orf 4	hypothetical protein	ERZ97957	*R. irregularis* (AMF)	-
Contig 50 - orf 2	hypothetical protein	ESA08063	*R. irregularis* (AMF)	Pkinase_Tyr
Contig 52 - orf 6	hypothetical protein	ESA21130	*R. irregularis* (AMF)	-
Contig 53 - orf 1	hypothetical protein	ESA03396	*R. irregularis* (AMF)	Pkinase_Tyr

- indicates no Pfam domain detected.


1 10 20 30 40 50 60 70 80 85
MRE_DNA ATGTTAGGGAAAGTAATTTTAATAG
MRE_aa M L G K V I L I
Ri_DNA CTAACGCAATTAATTACTTGAGCCTACTATTGGTGAAATGTGAACTTAACTTTCTCAAATATGTTA--ACCTCTACTTTATACTA
Ri_mRNA ATGTTA--ACCTCTACTTTATACTA
Ri_aa M L -- T S T L Y Y
MRE_DNA TAATAATTTTATTTTTGGTGTGAGTTATTAGTTGGTTTATTCCTTCACCAACTAAGAGTGATAAAAATACCGGGCAAATTAGTGA
MRE_aa V I I L F L V W V I S W F I P S P T K S D K N T G Q I S D
Ri_DNA TAAAGACGCTATGCTTGCTCCAGAGGTATAATTATAATATTTCTTATAAATTTTTGTATCGAAAGTATGTTCTTTAAATTATTTT
Ri_mRNA TAAAGACGCTATGCTTGCTCCAG
Ri_aa K D A M L A P
MRE_DNA TTGGGAATATCAGAAAAAACAAGCAAAACGGGAAGCAGAAGCAGAAAAGAATAAAGAATGATTTAAAGAGGGGAGATGGTTGACG
MRE_aa W E Y Q K K Q A K R E A E A E K N K E W F K E G R W L T
Ri_DNA TTTTTTTCACC----------CAATACAGGATTATGGACTCTTATACGATGATAGATGGTTGAAT
Ri_mRNA AGGATTATGGACTCTTATACGATGATAGATGGTTGAAT
Ri_aa E D Y G L L Y D D R W L N
MRE_DNA GATAAGGAAATTGATTGG
MRE_aa D K E I D W
Ri_DNA GACCGATGCGTAGATTTTTATTTGGAGTATGTAAGAAATCATTACTTGATTTTCATAAATATTTCTAAAAATAAAACATTTTTTT
Ri_mRNA GACCGATGCGTAGATTTTTATTTGGA
Ri_aa D R C V D F Y L E
MRE_DNA GCTACGGATAGATTATCTAAGGATAAAAGATTTAAAATCTTACCTGCTCACCAA
MRE_aa A T D R L S K D K R F K I L P A H Q
Ri_DNA TTCCTTTAAAAGATATCTTGAATACAATTATTCTATAAAGAGAGAATCCAAGGTTAAA---CCAAATCTTTTGCGAGCATCG---
Ri_mRNA ATATCTTGAATACAATTATTCTATAAAGAGAGAATCCAAGGTTAAA---CCAAATCTTTTGCGAGCATCG---
Ri_aa Y L E Y N Y S I K E E S K V K --- P N L L R A S
MRE_DNA TTTCATTATGTGCGAGAAGTAACTGATAAAAAAGAAACTG
MRE_aa F H Y V R E V T D K K E T
Ri_DNA ATGAGTTTTCTCATAACAAAT--ATTGAAGGTTTTTTTCATCCACTTTTTTTTTGTTCTAATATCATTTACCT
Ri_mRNA ATGAGTTTTCTCATAACAAAT--ATTG
Ri_aa M S F L I T N -- I
MRE_DNA GTTGAGAATTATCTT-TTAAAGAACTACTAAACCAAATAAACGAC
MRE_aa G W E L S -F K E L L N Q I N D
Ri_DNA ATTTTATTCTATTTACATATATTTCATTTATTTTTATAAAGATGCGAAATATCTTGCTAGCGCACTTCCAAAACAAATATTT---
Ri_mRNA AAGATGCGAAATATCTTGCTAGCGCACTTCCAAAACAAATATTT---
Ri_aa E D A K Y L A S A L P K Q I F ---
MRE_DNA CCTAGTAAAGAACTAATTTTTATCCCCATAAATAACCCCAATTTC----------CATTGGAGTTTGTTGG
MRE_aa P S K E L I F I P I N N P N F ---------- H W S L L
Ri_DNA ---TCTGCTGATATAATATTCATTCCAGTAAATAATAAACAGTCCTTAGAAAGTTTTATTGGAGGAAGTCATTGGTCTTTGCTAA
Ri_mRNA ---TCTGCTGATATAATATTCATTCCAGTAAATAATAAACAGTCCTTAGAAAGTTTTATTGGAGGAAGTCATTGGTCTTTGCTAA
Ri_aa --- S A D I I F I P V N N K Q S L E S F I G G S H W S L L
MRE_DNA TTTATGAAATCTCAACTAAAAAATTTTATCACTATGATACTCTCCAAGGGGCCAACGAT---AAT---TATATTAAACCTCTAGT
MRE_aa V Y E I S T K K F Y H Y D T L Q G A N D --- N --- Y I K P L V
Ri_DNA TTTACGTAAAACAGAATAATAAATTTTTATATTATGATA---------GTGCTAATAATATGAATTCTTATATT--GCTTATA--
Ri_mRNA TTTACGTAAAACAGAATAATAAATTTTTATATTATGATA---------GTGCTAATAATATGAATTCTTATATT--GCTTATA--
Ri_aa I Y V K Q N N K F L Y Y D ---------S A N N M N S Y I -- A Y --
MRE_DNA GAGAGAATTAGTGGAACAAATTCAAGCGGTAAGGAGTATCAAAGAAGATTATTTAAGGA----------
MRE_aa R E L V E Q I Q A V R S I K E D Y L R ----------
Ri_DNA -----AATTTGCAGATAAAATTTCCGGTGTTCTAGAAGTCAAAAGTAGGTATTAAAGAAATGAAAAATTATATTATGATTAGATT
Ri_mRNA -----AATTTGCAGATAAAATTTCCGGTGTTCTAGAAGTCAAAAG----------
Ri_aa -----K F A D K I S G V L E V K R----------
MRE_DNA GATATTTAATTAATAAACATGATCTAAGGCAAAATAATGGT---
MRE_aa R Y L I N K H D L R Q N N G ---
Ri_DNA ATGATATTTTTTTTAAAAAAAAAATTCCATTTTTAGGAAATAATATTGTCGTTGTAAAAACACCTCAA--CAAAAAAATGGTATG
Ri_mRNA AAATAATATTGTCGTTGTAAAAACACCTCAA--CAAAAAAATGGT---
Ri_aa N N I V V V K T P Q -- Q K N G ---

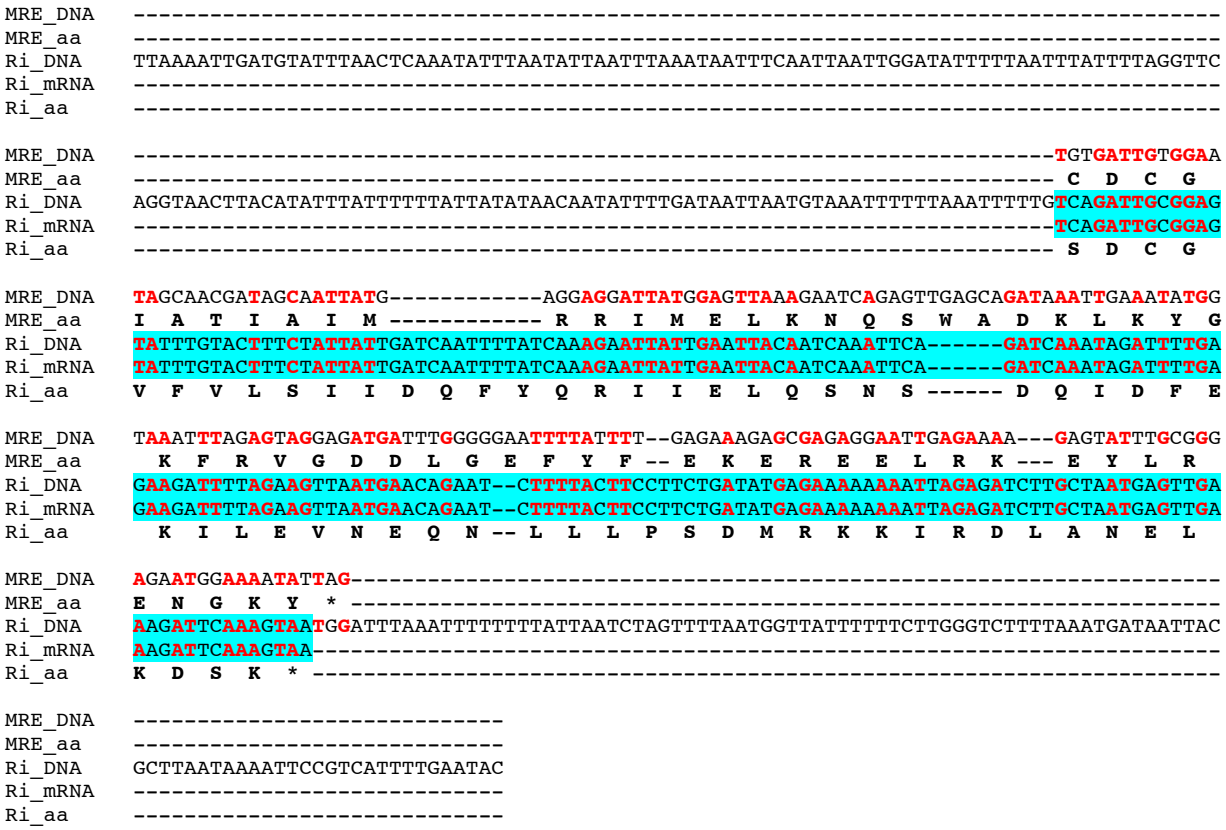


Figure S3.1. Sequence alignment of SUMO protease genes of AMF and MRE. This figure is the actual nucleotide alignment used to create Figure 4. MUSCLE was used for the sequence alignment (Edgar 2004). The gene from MRE-RV, contig 24, orf 13 is used as a representative of MRE, along with the SUMO protease sequence from *R. irregularis* (Ri) as a representative of AMF (GenBank Accessions ESA14994/KI282387). The blue boxes represent the exons of the Ri gene. The aligned MRE nucleotides that have identity to Ri sequences are in red and bold text. The amino acid sequences (aa) for both MRE and Ri sequences are also shown.

REFERENCES

Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y (2010). Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol* **18:** 132-139.

Allen JM, Light JE, Perotti MA, Braig HR, Reed DL (2009). Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet. *PLoS One* **4:** e4969.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Ashida H, Kim M, Sasakawa C (2014). Exploitation of the host ubiquitin system by human bacterial pathogens. *Nat Rev Microbiol* **12:** 399-413.

Bago B, Pfeffer PE, Shachar-Hill Y (2000). Carbon metabolism and transport in arbuscular mycorrhizas. *Plant Physiol* **124:** 949-958.

Bianciotto V, Lumini E, Bonfante P, Vandamme P (2003). '*Candidatus* Glomeribacter gigasporarum' gen. nov., sp. nov., an endosymbiont of arbuscular mycorrhizal fungi. *Int J Syst Evol Microbiol* **53:** 121-124.

Bierne H, Sabet C, Personnic N, Cossart P (2007). Internalins: a complex family of leucine-rich repeat-containing proteins in *Listeria monocytogenes*. *Microbes Infect* **9:** 1156-1166.

Bischof DF, Vilei EM, Frey J (2006). Genomic differences between type strain PG1 and field strains of *Mycoplasma mycoides* subsp. mycoides small-colony type. *Genomics* **88:** 633-641.

Brelsfoard C, Tsiamis G, Falchetto M, Gomulski LM, Telleria E, Alam U *et al* (2014). Presence of extensive Wolbachia symbiont insertions discovered in the genome of its host *Glossina morsitans* morsitans. *PLoS Negl Trop Dis* **8:** e2728.

Brinster S, Posteraro B, Bierne H, Alberti A, Makhzami S, Sanguinetti M *et al* (2007). Enterococcal leucine-rich repeat-containing protein involved in virulence and host inflammatory response. *Infect Immun* **75:** 4463-4471.

Canbäck B, Tamas I, Andersson SGE (2004). A phylogenomic study of endosymbiotic bacteria. *Molecular Biology and Evolution* **21:** 1110-1122.

Carle P, Saillard C, Carrere N, Carrere S, Duret S, Eveillard S *et al* (2010). Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important gene decay. *Appl Environ Microbiol* **76:** 3420-3426.

Cranenbrouck S, Voets L, Bivort C, Renard L, Strullu D, Declerck S (2005). Methodologies for *in vitro* cultivation of arbuscular mycorrhizal fungi with root organs. In: Declerck S, Fortin JA, Strullu D (eds). *In Vitro Culture of Mycorrhizas*. Springer Berlin Heidelberg. pp 341-375.

Da Lage JL, Feller G, Janecek S (2004). Horizontal gene transfer from Eukarya to bacteria and domain shuffling: the alpha-amylase model. *Cell Mol Life Sci* **61:** 97-109.

Das B, Martinez E, Midonet C, Barre FX (2013). Integrative mobile elements exploiting Xer recombination. *Trends Microbiol* **21:** 23-30.

Delaney NF, Balenger S, Bonneaud C, Marx CJ, Hill GE, Ferguson-Noel N *et al* (2012). Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*. *PLoS Genet* **8:** e1002511.

Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23:** 673-679.

den Bakker HC, Vankuren NW, Morton JB, Pawlowska TE (2010). Clonality and recombination in the life history of an asexual arbuscular mycorrhizal fungus. *Mol Biol Evol* **27:** 2474-2486.

Desiro A, Naumann M, Epis S, Novero M, Bandi C, Genre A *et al* (2012). Mollicutes-related endobacteria thrive inside liverwort-associated arbuscular mycorrhizal fungi. *Environ Microbiol*.

Desiro A, Salvioli A, Ngonkeu EL, Mondo SJ, Epis S, Faccio A *et al* (2014). Detection of a novel intracellular microbiome hosted in arbuscular mycorrhizal fungi. *ISME J* **8:** 257-270.

Dupuy V, Manso-Silvan L, Barbe V, Thebault P, Dordet-Frisoni E, Citti C *et al* (2012). Evolutionary history of contagious bovine pleuropneumonia using next generation sequencing of *Mycoplasma mycoides* subsp. mycoides "Small Colony". *PLoS One* **7:** e46821.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792-1797.

Everett RD, Boutell C, Hale BG (2013). Interplay between viruses and host sumoylation pathways. *Nat Rev Microbiol* **11:** 400-411.

Fellbaum CR, Gachomo EW, Beesetty Y, Choudhari S, Strahan GD, Pfeffer PE *et al* (2012). Carbon availability triggers fungal nitrogen uptake and transport in arbuscular mycorrhizal symbiosis. *Proc Natl Acad Sci U S A* **109:** 2666-2671.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR *et al* (2014). Pfam: the protein families database. *Nucleic Acids Res* **42:** D222-230.

Garcia-Vallve S, Romeu A, Palau J (2000). Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res* **10:** 1719-1725.

Gardebrecht A, Markert S, Sievert SM, Felbeck H, Thurmer A, Albrecht D *et al* (2012). Physiological homogeneity among the endosymbionts of *Riftia pachyptila* and *Tevnia jerichonana* revealed by proteogenomics. *ISME J* **6:** 766-776.

Gasiunas G, Sinkunas T, Siksnys V (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. *Cell Mol Life Sci* **71:** 449-465.

Gianinazzi S, Gollotte A, Binet MN, van Tuinen D, Redecker D, Wipf D (2010). Agroecology: the key role of arbuscular mycorrhizas in ecosystem services. *Mycorrhiza* **20:** 519-530.

Goecks J, Nekrutenko A, Taylor J, Galaxy T (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11:** R86.

Gomez-Valero L, Bonora MN, Gribaldo S, Buchrieser C (2013). Interdomain horizontal gene transfer shaped the genomes of *Legionella pneumophila* and *Legionella longbeachae*. *Lateral Gene Transfer in Evolution*. Springer New York. pp 199-219.

Hartfield M, Otto SP, Keightley PD (2012). The maintenance of obligate sex in finite, structured populations subject to recurrent beneficial and deleterious mutation. *Evolution* **66:** 3658-3669.

Helber N, Wippel K, Sauer N, Schaarschmidt S, Hause B, Requena N (2011). A versatile monosaccharide transporter that operates in the arbuscular mycorrhizal fungus *Glomus* sp is crucial for the symbiotic relationship with plants. *Plant Cell* **23:** 3812-3823.

Hotson A, Chosed R, Shu H, Orth K, Mudgett MB (2003). Xanthomonas type III effector XopD targets SUMO-conjugated proteins in planta. *Mol Microbiol* **50:** 377-389.

Ishida K, Sekizuka T, Hayashida K, Matsuo J, Takeuchi F, Kuroda M *et al* (2014). Amoebal Endosymbiont Neochlamydia Genome Sequence Illuminates the Bacterial Role in the Defense of the Host Amoebae against *Legionella pneumophila*. *PLoS One* **9:** e95166.

Ishii Y, Kakizawa S, Oshima K (2013). New *ex vivo* reporter assay system reveals that sigma factors of an unculturable pathogen control gene regulation involved in the host switching between insects and plants. *Microbiologyopen* **2:** 553-565.

Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ (2010). Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. *Science* **329:** 212-215.

Jubelin G, Taieb F, Duda DM, Hsu Y, Samba-Louaka A, Nobe R *et al* (2010). Pathogenic bacteria target NEDD8-conjugated cullins to hijack host-cell signaling pathways. *PLoS Pathog* **6:** e1001128.

Juhasz J, Kertesz-Farkas A, Szabo D, Pongor S (2014). Emergence of collective territorial defense in bacterial communities: horizontal gene transfer can stabilize microbiomes. *PLoS One* **9:** e95511.

Kelman Z, O'Donnell M (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu Rev Biochem* **64:** 171-200.

Ku C, Lo WS, Chen LL, Kuo CH (2013). Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in spiroplasma. *Genome Biol Evol* **5:** 1151-1164.

Leclercq S, Giraud I, Cordaux R (2011). Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. *Mol Biol Evol* **28:** 685-697.

Li Y, Zheng H, Liu Y, Jiang Y, Xin J, Chen W *et al* (2011). The complete genome sequence of *Mycoplasma bovis* strain Hubei-1. *PLoS One* **6:** e20999.

Lowe TM, Eddy SR (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955-964.

Luo C, Tsementzi D, Kyrpides NC, Konstantinidis KT (2012). Individual genome assembly from complex community short-read metagenomic datasets. *ISME J* **6:** 898-901.

Macdonald RM, Chandler MR (1981). Bacterium-like organelles in the vesicular-arbuscular mycorrhizal fungus *Glomus caledonius*. *New Phytol* **89:** 241-246.

Maniloff J (2002). Phylogeny and evolution. In: Razin S, Herrmann R (eds). *Molecular Biology and Pathogenicity of Mycoplasmas*. Kluwer Academic/Plenum Publishers: New York, NY.

Marenda MS (2014). Genomic Mosaics. In: Browning GF, Citti C (eds). *Mollicutes Molecular Biology and Pathogenesis*. Caister Academic Press: Norfolk, UK. pp 15-54.

Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y *et al* (2012). IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* **40:** D115-122.

Mondo SJ, Toomer KH, Morton JB, Lekberg Y, Pawlowska TE (2012). Evolutionary stability in a 400-million-year-old heritable facultative mutualism. *Evolution* **66:** 2564-2576.

Moran NA, McCutcheon JP, Nakabachi A (2008). Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42:** 165-190.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35:** W182-185.

Morris JJ, Lenski RE, Zinser ER (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *MBio* **3**.

Mosse B (1970). Honey-coloured, sessile Endogone spores: II. Changes in fine structure during spore development. *Arch Mikrobiol* **74:** 129-145.

Mrazek J (2006). Analysis of distribution indicates diverse functions of simple sequence repeats in Mycoplasma genomes. *Mol Biol Evol* **23:** 1370-1385.

Muller HJ (1964). The Relation of Recombination to Mutational Advance. *Mutat Res* **106:** 2-9.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012). MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. *Nucleic Acids Res*.

Naumann M, Schussler A, Bonfante P (2010). The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. *ISME J* **4:** 862-871.

Nikolaidis N, Doran N, Cosgrove DJ (2014). Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion. *Mol Biol Evol* **31:** 376-386.

O'Fallon B (2008). Population structure, levels of selection, and the evolution of intracellular symbionts. *Evolution* **62:** 361-373.

Ochman H, Elwyn S, Moran NA (1999). Calibrating bacterial evolution. *Proc Natl Acad Sci U S A* **96:** 12638-12643.

Osburne MS, Holmbeck BM, Coe A, Chisholm SW (2011). The spontaneous mutation frequencies of *Prochlorococcus* strains are commensurate with those of other bacteria. *Environ Microbiol Rep* **3:** 744-749.

Pan X, Luhrmann A, Satoh A, Laskowski-Arce MA, Roy CR (2008). Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320:** 1651-1654.

Paoletti M, Clave C (2007). The fungus-specific HET domain mediates programmed cell death in *Podospora anserina*. *Eukaryot Cell* **6:** 2001-2008.

Parniske M (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Microbiol* **6:** 763-775.

Rabut G, Peter M (2008). Function and regulation of protein neddylation. 'Protein modifications: beyond the usual suspects' review series. *EMBO Rep* **9:** 969-976.

Redecker D, Kodner R, Graham LE (2000). Glomalean fungi from the Ordovician. *Science* **289:** 1920-1921.

Reuber TL, Ausubel FM (1996). Isolation of Arabidopsis genes that differentiate between resistance responses mediated by the RPS2 and RPM1 disease resistance genes. *Plant Cell* **8:** 241-249.

Ribet D, Hamon M, Gouin E, Nahori MA, Impens F, Neyret-Kahn H *et al* (2010). *Listeria monocytogenes* impairs SUMOylation for efficient infection. *Nature* **464:** 1192-1195.

Richards AM, Von Dwingelo JE, Price CT, Abu Kwaik Y (2013). Cellular microbiology and molecular ecology of Legionella-amoeba interaction. *Virulence* **4:** 307-314.

Ron Y, Flitman-Tene R, Dybvig K, Yogev D (2002). Identification and characterization of a site-specific tyrosine recombinase within the variable loci of *Mycoplasma bovis*, *Mycoplasma pulmonis* and *Mycoplasma agalactiae*. *Gene* **292:** 205-211.

Ronquist F, Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572-1574.

Sampson TR, Weiss DS (2013). Alternative roles for CRISPR/Cas systems in bacterial pathogenesis. *PLoS Pathog* **9:** e1003621.

Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T *et al* (2010). The genome of the amoeba symbiont "*Candidatus* Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* **192:** 1045-1057.

Sirand-Pugnet P, Citti C, Barre A, Blanchard A (2007a). Evolution of mollicutes: down a bumpy road with twists and turns. *Res Microbiol* **158:** 754-766.

Sirand-Pugnet P, Lartigue C, Marenda M, Jacob D, Barre A, Barbe V *et al* (2007b). Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. *PLoS Genet* **3:** e75.

Sloan DB, Moran NA (2013). The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biol Evol* **5:** 783-793.

Smith SE, Read DJ (2008). *Mycorrhizal Symbiosis*, Third edn. Academic Press: New York.

Sogaard IZ, Boesen T, Mygind T, Melkova R, Birkelund S, Christiansen G *et al* (2002). Recombination in *Mycoplasma hominis*. *Infect Genet Evol* **1:** 277-285.

Szczepanek SM, Tulman ER, Gorton TS, Liao X, Lu Z, Zinski J *et al* (2010). Comparative genomic analyses of attenuated strains of Mycoplasma gallisepticum. *Infect Immun* **78:** 1760-1771.

Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ *et al* (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296:** 2376-2379.

Thompson CC, Vieira NM, Vicente AC, Thompson FL (2011). Towards a genome based taxonomy of Mycoplasmas. *Infect Genet Evol* **11:** 1798-1804.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R *et al* (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences*.

Toomer KH, Chen X, Naito M, Mondo SJ, den Bakker HC, VanKuren NW *et al* (2014). Are the *Mycoplasma*-related endobacteria ancient parasites of arbuscular mycorrhizal fungi? *In Review*.

Wang B, Qiu YL (2006). Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16:** 299-363.

Wang Z, Li X (2009). IAN/GIMAPs are conserved and novel regulators in vertebrates and angiosperm plants. *Plant Signal Behav* **4:** 165-167.

Westra ER, Buckling A, Fineran PC (2014). CRISPR-Cas systems: beyond adaptive immunity. *Nat Rev Microbiol* **12:** 317-326.

Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN *et al* (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462:** 1056-1060.

Xiao L, Paralanov V, Glass JI, Duffy LB, Robertson JA, Cassell GH *et al* (2011). Extensive horizontal gene transfer in ureaplasmas from humans questions the utility of serotyping for diagnostic purposes. *J Clin Microbiol* **49:** 2818-2826.

Zhang M, Pereira ESM, De Mares MC, van Elsas J (2014). The mycosphere constitutes an arena for horizontal gene transfer with strong evolutionary implications for bacterial-fungal interactions. *FEMS Microbiol Ecol*.

CHAPTER 4

DEFYING MULLER'S RATCHET: HERITABLE ENDOBACTERIA ESCAPE EXTINCTION THROUGH RECOMBINATION AND GENOME PLASTICITY.[1]

4.1 - ABSTRACT

Heritable endobacteria, which are transmitted from one host generation to the next, are subjected to evolutionary forces that are distinctly different from those experienced by free-living bacteria. In particular, they suffer consequences of Muller's ratchet, an evolutionary mechanism that leads to extinction of small asexual populations due to continuous fixation of slightly deleterious mutations combined with the random loss of the most fit individuals, which cannot be recreated in the absence of recombination. Muller's ratchet progression has been observed empirically in many heritable endobacteria with degenerate eroded genomes. Mollicutes/mycoplasma-related endobacteria (MRE) are heritable endobacteria that reside in the cytoplasm of all major lineages of arbuscular mycorrhizal fungi (Glomeromycota). Previous studies revealed that MRE maintain unusually diverse populations inside their hosts. Here we show that these endobacteria are ancient fungal associates and they defy Muller's ratchet through adaptive retention of recombination to maintain plastic genomes and high population diversity. MRE are, to our knowledge, the first ancient heritable endosymbionts that engage in bacterial sex and do not suffer the consequences of Muller's ratchet typical for vertically transmitted bacteria. We propose that other endobacteria may be also capable of adapting similar defenses against Muller's ratchet.

[1]The results of this study will be submitted to *mBio* as an Opinions & Hypothesis article, and is written according to their guidelines. All experiments were performed by and manuscript written by M.N.

4.2 - RESULTS/DISCUSSION

Molecular Evolution Patterns Of Heritable Endobacteria

Heritable endobacteria live within eukaryotic host cells, and are transmitted from one host generation to the next. They occupy cytoplasmic niches in a variety of hosts, including insects, nematodes, and fungi (Hoffmeister and Martin 2003, Naumann et al 2010). The associations of bacteria with eukaryotic hosts often result in evolutionary transitions that generate novel metabolic capabilities that each partner living alone could not achieve. The extent of interdependence between the host and endobacteria can vary based on factors such as age of association and degree of co-evolution, with some associations, such as mitochondria and chloroplasts of eukaryotic cells, achieving ultimate integration (Moya et al 2008).

Vertically transmitted endobacteria undergo unique molecular evolution, resulting in genome structures generally not seen in their free-living relatives. Endosymbiont genomes are characterized by extreme reduction, often bias towards adenine and thymine nucleotide composition, and accelerated sequence evolution (Moran et al 2008). These features are consequences of living in the cytoplasmic niche of a eukaryotic host that is unlike a free-living environment. The intracellular environment offers not only protection and absence of competition, but a metabolically rich milieu where many genes in the endobacteria's genome become redundant or unnecessary (Moya et al 2008). Furthermore, the exclusively vertical transmission has several consequences that affect endosymbiont population structure. In particular, only a portion of the endosymbiont population is passed to the next generation of the host, creating transmission bottlenecks. Furthermore, endosymbiont populations are clonal and confined to the host lineages. All these factors contribute to reduction of endosymbiont effective population sizes relative to free-living bacteria. Such population dynamics create a situation

where purifying selection is weak, resulting in the accumulation and fixation of slightly deleterious mutations in the endosymbiont genome. These mutations can lead to the inactivation of the affected genes, followed by their deletions, ultimately creating a reduced and degenerate genome (McCutcheon and Moran 2012, Nilsson et al 2005).

The genes that are ultimately lost in reduced endobacterial genomes come from all functional categories, most notably genes involved in metabolism, cell envelope biosynthesis, transcriptional regulation, and DNA repair and recombination. The genes that are retained, on the other hand, are involved in essential cellular functions, such as DNA replication, RNA transcription, and protein translation (McCutcheon and Moran 2012). However, even within these essential categories, only the minimal genes required for basic functionality are maintained, with most endobacteria displaying the loss of accessory subunits otherwise present in these systems. Chaperone proteins involved in the stability of protein folding are also consistently maintained in endobacteria, likely rendered essential as deleterious mutations accumulate throughout the genome (Moya et al 2008).

Muller's Ratchet And Population Extinction

The absence of recombination, and frequent population bottlenecks make heritable endobacterial populations vulnerable to extinction. As slightly deleterious mutations continuously fix into the population, the fittest genotypes will be progressively lost by chance. This continual decrease in the mean fitness of a population was described by Hermann Muller, and is known as Muller's ratchet (Crow 2005, Muller 1964). The loss of the most fit class of genotypes advances the ratchet towards population extinction, which cannot be reverted in the absence of recombination. Muller's ratchet is also believed to be the driving force behind the degenerate and eroded

genomes of heritable endobacteria (Moran 1996, Pettersson and Berg 2007). While recombination is the ultimate mechanism that can decelerate advance of the ratchet, several other factors are also known to slow it down, including compensatory evolution (Poon and Otto 2000) and host-level selection (Pettersson and Berg 2007).

Finally, Muller's ratchet and transmission bottlenecks experienced by heritable endobacteria contribute to reduction of population genetic diversity. As a consequence, ancient endobacteria are characterized by genetic homogeneity within a host (Tamas et al 2002).

Unique Genome Evolution In Fungal Endobacteria

Recently, we reported on a metagenomic study of a unique group of heritable endobacteria, the Mollicutes/mycoplasma-related endobacteria (MRE) (Chapter 3). MRE are distributed world-wide, and reside in the cytoplasm of all major lineages of arbuscular mycorrhizal fungi (AMF). AMF are ancient asexual soil fungi that themselves are obligate symbionts of majority of terrestrial plants (Naumann et al 2010, Parniske 2008, Toomer et al 2014). As the name suggests, MRE are members of the Mollicutes, a group characterized by minimal genome and lack of a peptidoglycan cell wall, despite their phylogenetic affinity to Gram positive bacteria. Interestingly, MRE populations have unusually high diversity levels within their hosts, with extremely plastic genomes suggested by extensive chromosomal rearrangements (Chapter 3).

Since the age of association between endobacteria and their hosts will determine the amount of molecular evolution expected in the bacterial population, we first estimated the date of origin of the MRE lineage (Figure 4.1). We used BEAST (Bayesian Evolutionary Analysis Sampling Trees) (Bouckaert et al 2014) with 16S rRNA sequences and two calibration points: (1) 190 million years for the divergence of *Mycoplasma mobile* and *M. pulmonis*, and (2) 450

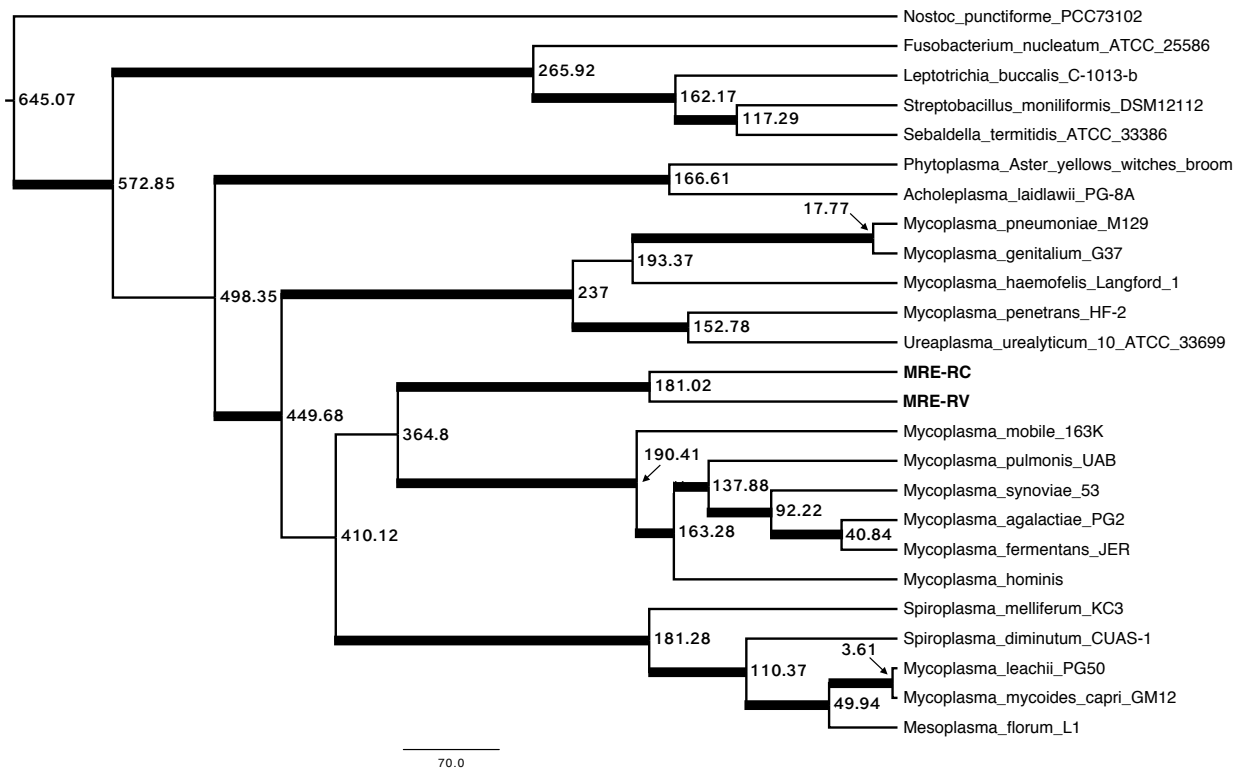


Figure 4.1. Divergence dating based on 16S rRNA gene sequences indicates MRE are an ancient lineage. Numbers at nodes indicate the age of divergence, in millions of years. Posterior probabilities greater than 0.90 are indicated by thickened branches.

million years for the divergence of the mycoplasmas from the rest of the Mollicutes (Jaffe et al 2004, Maniloff 2002). BEAST was run using the GTR+I+Γ model for 1,000,000 generations, with the lognormal relaxed clock model. All sequences are obtained from Chapter 2 and 3. This analysis indicates that MRE share ancestry with the pneumoniae group of the Mycoplasmatales and have diverged from their sister clade364.8 million years ago (with 95% highest posterior density interval of 302.6 and 418.5). If we assume that this date also marks a transition from animal to fungal hosts then the association of MRE with AMF is an ancient event. While transmission of present-day animal-infecting mycoplasmas is predominantly horizontal, MRE are, for the most part, transmitted vertically in asexual fungi, and this host switch must have also generated novel selection pressures on the MRE transmission mode.

The metagenomic data of the MRE revealed the presence of various DNA recombination genes (Chapter 3), despite their losses in all other heritable endobacteria (McCutcheon and Moran 2012). This is an unusual feature in endobacteria, as genes that are not essential for basic cellular processes are inevitably lost and removed from endobacterial genomes due to the Muller's ratchet. To determine if these DNA recombination genes are functional, we looked for evidence of recombination in the MRE genomes using GARD (Genetic Algorithm for Recombination Detection) (Kosakovsky Pond et al 2006). Using clonally isolated ~1100 base partial sequence of the 16S rRNA gene from 4 MRE populations, we found evidence of recombination in two breakpoints, at the aligned positions 276 and 476, with an AIC_c score of 8459.13 compared to the score of 8956.05 for no breakpoints. The Kishino-Hasegawa test further confirmed significant topological incongruence between the segments at each breakpoint with a $P < 0.001$ (Figure 4.2). These results indicate that recombination is occurring in the MRE

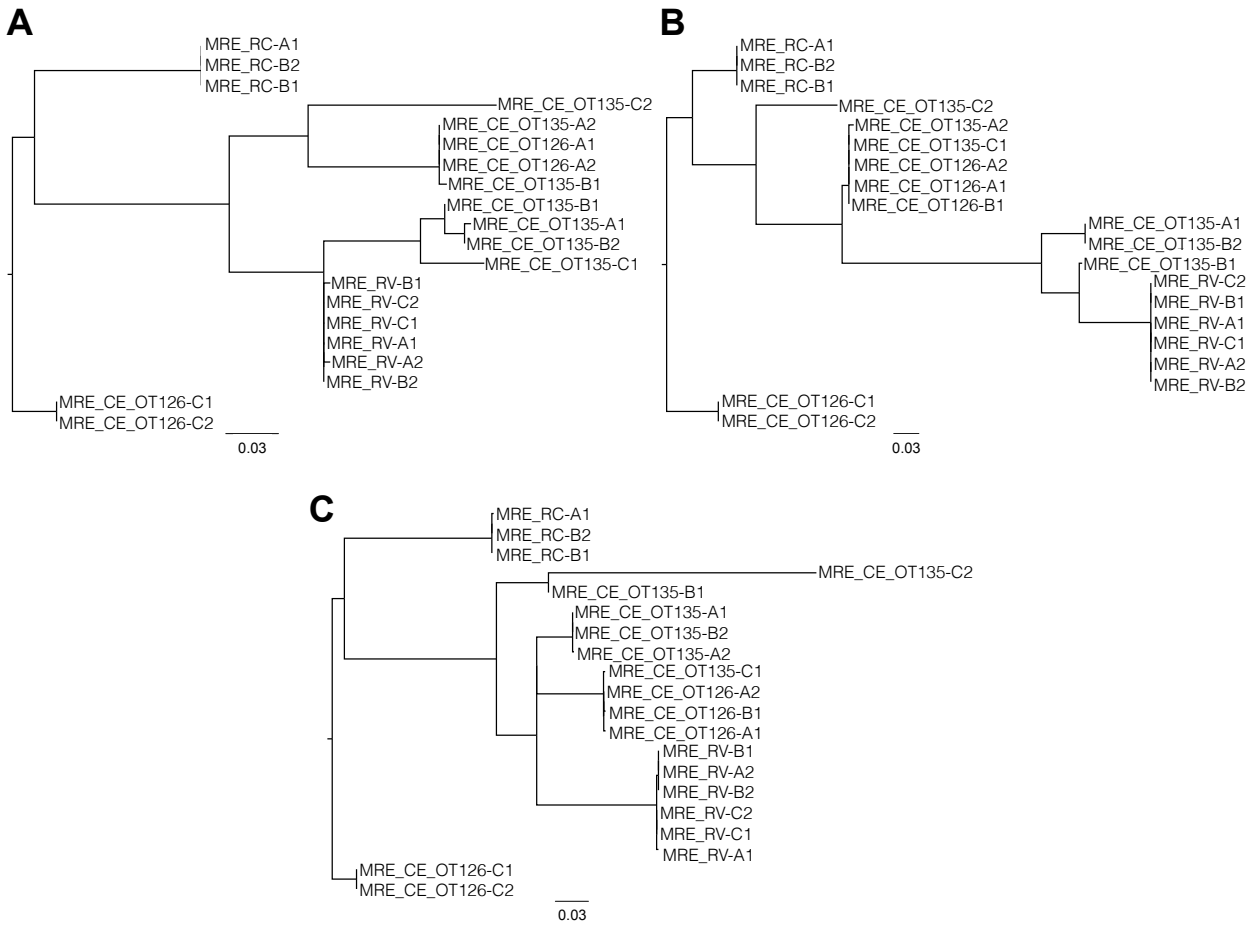


Figure 4.2. Evidence of recombination within MRE populations. Phylogeny reconstructions for 16S rRNA gene alignment segments separated by recombination points at: (A) positions 1-276, (B) positions 277-476 and (C) positions 477-1117. MRE-RC are MRE populations from the host species *Rhizophagus clarus* NB112A, MRE-RV are MRE populations from the host species *Racocetra verrucosa* VA103A, MRE-CE are MRE populations from the host species *Claroideoglomus etunicatum*, from population accessions OT135 and OT126. Letter designations are individual spores, and final number designations are 16S rRNA sequence clones.

genomes, with individuals with functional recombination genes being selected for to maintain a highly diverse population within the fungal cytoplasm.

Adaptive Maintenance Of Recombination To Overcome Muller's Ratchet

We equate the adaptive maintenance of recombination machinery in the MRE genome to the evolution of sex in eukaryotic systems; despite its inherent costs, sex is maintained in the majority of eukaryotes, with theories of its role ranging from the creation of novel genotypes to resist infection (Red Queen Hypothesis), to fixation of multiple novel advantageous mutants (Fisher-Muller Hypothesis), to the prevention of population extinction via Muller's ratchet in finite populations (Hartfield and Keightley 2012). Though the full mechanism is unclear, MRE are capable of genomic recombination, likely through the uptake of DNA from dead MRE cells in the same host cytoplasmic niche. DNA uptake and recombination can combine MRE genomes of intermediate fitness to create an individual cell with higher fitness. Being subjected to predominantly vertical transmission, the MRE genomes inevitably continue to erode through Muller's ratchet, but may purge the deleterious mutations with subsequent DNA uptake and recombination (Figure 4.3). Furthermore, theoretical modeling indicates that horizontal gene transfer of foreign DNA in bacterial populations can prevent the operation of Muller's ratchet, even if the foreign genes contain more deleterious mutations than the recipient cells (Takeuchi et al 2014). If uptake and recombination of foreign DNA is frequent enough, a diverse bacterial population can resist Muller's ratchet better than a genetically homogenous population of the same size. It is clear that maintaining diversity through recombination can prevent the further erosion and potential extinction of endobacterial populations. It is unclear how MRE are able to prevent their recombination machinery from being eroded, as seen in the majority of

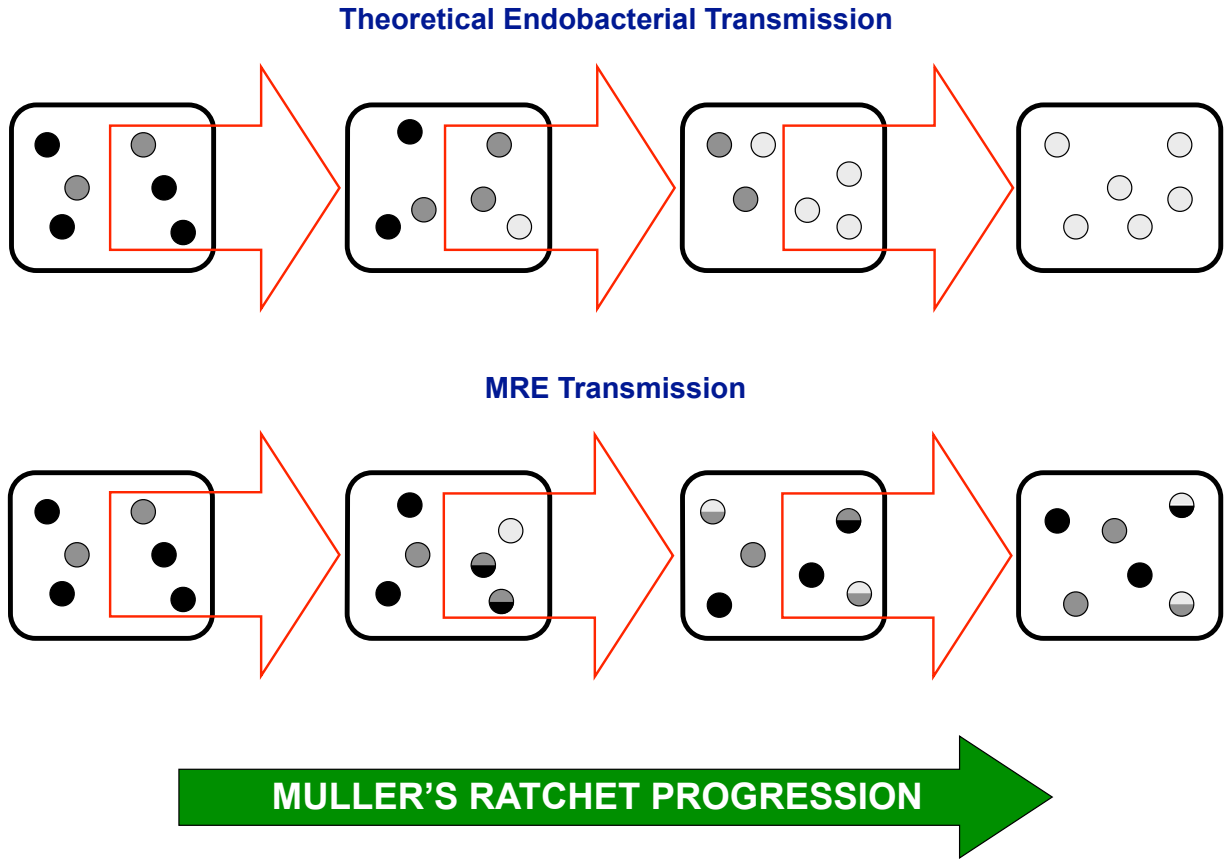


Figure 4.3. Cartoon depicting MRE host transmission and Muller's ratchet progression compared to theoretical progression. According to Muller's ratchet, due to host transmission bottlenecks and fixation of deleterious mutations, endobacterial populations will degenerately progress to a relatively homogeneous population of unfit individuals. MRE escape Muller's ratchet by maintenance of recombination to purge deleterious mutations and maintain a genetically diverse population. Black outlined rectangles represents host cells, with red arrows indicating transmission bottlenecks. Endobacteria are represented as small circles, with darker shading depicting the most fit individuals to lighter shading depicting the least fit individuals. Circles with two-tone colour depict recombination events in MRE.

endobacteria. This may be a general feature of the genome structure in all Mollicutes. As all members of the Mollicutes are characterized by reduced genomes, the MRE already harboured a minimal genome when they became associated with the fungal hosts. Due to their limited gene set, recombination genes may be more essential to the Mollicutes than to other endobacteria, subjecting these genes to stronger selective pressures than in other endobacteria. Furthermore, despite their minimal genome, other *Mycoplasma* species seem to have retained their sexual competence, and ability for horizontal gene transfer in mycoplasmas sharing the same niche (Sirand-Pugnet et al 2007). Taken together, MRE have been able to adapt a novel way of decelerating Muller's ratchet through bacterial sex, a phenomenon likely aided by their ancestry and relatedness to the Mollicutes class.

Non-Mollicute Endobacteria

It has been assumed for many years that evolution of heritable endobacteria follows a strict evolutionary path: early stages are characterized by an overabundance of mobile elements, pseudogenes, and multiple genomic rearrangements, while latter stages, or ancient endobacteria, are characterized by the lack of mobile elements, loss of most pseudogenes, and a relatively stable yet continuously reducing genome (McCutcheon and Moran 2012, Ochman and Davalos 2006). Recent studies and the increased sequencing of endobacterial genomes have revealed that this may not be the case. Genomes of *Wolbachia*, reproductive parasites of insects, revealed that despite their long-term association with insect hosts and reduced genomes, these endobacteria still harbour an over-abundance of mobile elements, and extensive genomic insertions (Brelsfoard et al 2014, Leclercq et al 2011). Analysis on the louse endosymbiont, *Candidatus* Riesia sp., revealed a reduction in mutation accumulation, leading to a slowdown of the effects

of Muller's ratchet (Allen et al 2009). Extensive genome rearrangement have also been documented in the genomes of *Portiera*, endosymbionts of whiteflies, likely due to recombination events across identical repeats found in the bacterial genome (Sloan and Moran 2013).

4.3 - CONCLUSIONS

Our data on the novel endobacteria of arbuscular mycorrhizal fungi revealed active maintenance of recombination genes in the population, leading to genetic diversity, despite the ancient age of the association. This allows for the prevention of Muller's ratchet (Figure 4.3), a luxury not possible in other endobacteria that have lost their recombinant abilities through genome erosion.

It is clear that genome evolution in endobacteria are influenced by multiple factors, and thus the trajectory of the endobacterial population may not be as simple and predictable as once believed. The evolution of endobacterial populations not only has profound effects on the population itself, but also for the host species that harbour them. The MRE are unique endobacteria that have adapted for bacterial sex, to prevent the effects of Muller's ratchet, not unlike higher eukaryotic species.

4.4 - ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation grant DEB-0918880 and by the Cornell University Agricultural Experiment Station federal formula funds, Project NYC-153434, received from the National Institutes for Food and Agriculture, U.S. Department of Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the U.S. Department of Agriculture.

REFERENCES

Allen JM, Light JE, Perotti MA, Braig HR, Reed DL (2009). Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet. *PLoS One* **4:** e4969.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C, Xie D *et al* (2014). BEAST2: A software platform for bayesian evolutionary analysis. *PLoS Comput Biol* **10:** e1003537.

Brelsfoard C, Tsiamis G, Falchetto M, Gomulski LM, Telleria E, Alam U *et al* (2014). Presence of extensive Wolbachia symbiont insertions discovered in the genome of its host Glossina morsitans morsitans. *PLoS Negl Trop Dis* **8:** e2728.

Crow JF (2005). Timeline: Hermann Joseph Muller, evolutionist. *Nat Rev Genet* **6:** 941-945.

Hartfield M, Keightley PD (2012). Current hypotheses for the evolution of sex and recombination. *Integr Zool* **7:** 192-209.

Hoffmeister M, Martin W (2003). Interspecific evolution: microbial symbiosis, endosymbiosis and gene transfer. *Environ Microbiol* **5:** 641-649.

Jaffe JD, Stange-Thomann N, Smith C, DeCaprio D, Fisher S, Butler J *et al* (2004). The complete genome and proteome of Mycoplasma mobile. *Genome Res* **14:** 1447-1461.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006). GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22:** 3096-3098.

Leclercq S, Giraud I, Cordaux R (2011). Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. *Mol Biol Evol* **28:** 685-697.

Maniloff J (2002). Phylogeny and Evolution. In: Razin S, Herrmann R (eds). *Molecular Biology and Pathogenicity of Mycoplasmas*. Springer US. pp 31-43.

McCutcheon JP, Moran NA (2012). Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10:** 13-26.

Moran NA (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* **93:** 2873-2878.

Moran NA, McCutcheon JP, Nakabachi A (2008). Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42:** 165-190.

Moya A, Pereto J, Gil R, Latorre A (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* **9:** 218-229.

Muller HJ (1964). The Relation of Recombination to Mutational Advance. *Mutat Res* **106:** 2-9.

Naumann M, Schussler A, Bonfante P (2010). The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. *ISME J* **4:** 862-871.

Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JC, Andersson DI (2005). Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci U S A* **102:** 12112-12116.

Ochman H, Davalos LM (2006). The nature and dynamics of bacterial genomes. *Science* **311:** 1730-1733.

Parniske M (2008). Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Microbiol* **6:** 763-775.

Pettersson ME, Berg OG (2007). Muller's ratchet in symbiont populations. *Genetica* **130:** 199-211.

Poon A, Otto SP (2000). Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution* **54:** 1467-1479.

Sirand-Pugnet P, Lartigue C, Marenda M, Jacob D, Barre A, Barbe V *et al* (2007). Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. *PLoS Genet* **3:** e75.

Sloan DB, Moran NA (2013). The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biol Evol* **5:** 783-793.

Takeuchi N, Kaneko K, Koonin EV (2014). Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. *G3 (Bethesda)* **4:** 325-339.

Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ *et al* (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296:** 2376-2379.

Toomer KH, Chen X, Naito M, Mondo SJ, den Bakker HC, VanKuren NW *et al* (2014). Are the *Mycoplasma*-related endobacteria ancient parasites of arbuscular mycorrhizal fungi? *In Review*.

CHAPTER 5

CONCLUSIONS & PERSPECTIVES

This dissertation research provided the first biological investigation on a novel group of endobacteria of AMF. AMF are not only important members of the rhizosphere that contribute significantly to global nutrient cycling, but their potential benefit as a replacement for non-renewable fertilization systems in agriculture points to the need for better understanding of this fungal group. As MRE are associated with all major lineages of AMF worldwide, their role in AMF biology was investigated.

In the first project, MRE were officially classified and given taxonomic placement in bacterial phylogeny. With the erection of a novel family, along with a genus and species name for the MRE of *R. clarus*, we provided the first reference point for others who are studying MRE populations. Furthermore, their taxonomic placement provided the support needed for evolutionary investigations of this group of endobacteria.

In the second project, MRE population metagenomes were constructed and analyzed to determine the biological capabilities of MRE, and their potential role in AMF hosts. Our data revealed that MRE genomes are highly plastic, with evidence of extensive chromosomal rearrangements. Furthermore, the MRE genomes encode for genes that are important for associations with eukaryotic hosts, likely affecting AMF biology. Finally, MRE genomes contain evidence of multiple horizontal gene transfers from the AMF host; some of these genes are involved in important eukaryotic cellular functions, and are likely being used for manipulation of the host cell. Overall, our findings provide the first glimpse at MRE biological capabilities, and the extent of AMF-MRE coevolution. This study is not only important for

furthering MRE-AMF research, but also understanding Mollicutes biology, and fungal-bacterial interactions.

The third project provided evidence for the unusual evolution of MRE. As endobacteria are under the pressure of unique evolutionary forces, their molecular evolution is expected to follow the theoretical path of Muller's ratchet. Our results show that MRE are able to defy Muller's ratchet through their ability to prevent degeneration of recombination genes. This in turn provides MRE the mechanism to maintain genetic diversity in the population, and prevent their population extinction. As more sequencing data for endobacterial species are being gathered, more evidence is being uncovered that show the adaptive ability of bacterial species to overcome theoretical expectations of population extinction. We provided the first evidence of an endobacterial species adapting to its surroundings by maintenance of bacterial sex to prevent the progression of Muller's ratchet.

Overall, the research project outlined in this dissertation has provided novel data on MRE-AMF biology, and contributed to the fields of microbiology, mycology, and molecular evolution.

Future work on MRE-AMF biology is still required to fully understand their association. Due to the lack of genetic manipulation techniques in either organisms and cultivation strategies for the MRE, the techniques available to investigate their biology are currently still limited. However, the first genome sequence of the endobacterial-free AMF species *R. irregularis* recently became available, providing new information on the biology of AMF. Future work should be focused on retrieving transcriptomes of AMF species that harbour MRE, to complement the biological data collected in this study. Furthermore, increasing MRE population

samples to other AMF host species and geographical locations will provide important information for comparative genomics and evolution of the MRE groups.

APPENDIX A

ATTEMPTS TO CREATE AN ISOGENIC AMF LINE WITH AND WITHOUT MRE

A.1 - RESULTS

In order to complement our genomic data, we attempted to create isogenic AMF lines that both harboured and lacked MRE. If successful, the two AMF lines would provide functional data on the roles of MRE in the AMF host. Two strategies were employed: curing an AMF species that harboured MRE (MRE+), and infecting an AMF species that was endobacteria-free (MRE−).

Curing MRE+ AMF of Their Endobacteria

The first attempt at curing MRE+ AMF of their endobacteria was through the use of antibiotics. The AMF species *R. clarus* was used for curing due to their ability to be cultivated *in vitro*. The first antibiotic used was Tylosin tartrate (Sigma), which contains a mixture of Tylosin A, B, C, and D tartrates, and is commonly used against various Mycoplasma species. Tylosin was added to a final concentration of 10 μg mL^{-1} directly on 1 week old sub-cultures of *R. clarus,* and freshly sub-cultured *R. clarus* spores. AMF grown directly within the media, and AMF growing in a single plane on cellophane was used. Tylosin was re-administered every 3 days for 1 month, after which the cultures were left to grow. New *R. clarus* spores in the cultures were tested for MRE presence using PCR, and MRE were detected in every case.

The above experiment was also repeated using the antibiotic Plasmocin™ (InvivoGen), an antimycoplasma agent, at a final concentration of 25 μg mL^{-1}. New *R. clarus* spores in the cultures were tested for MRE presence using PCR, and MRE were detected in every case.

The curing of *R. clarus* MRE+ AMF was also attempted using a serial dilution technique, where only a single spore is used as the primary inoculum for serial sub-culturing of the AMF, in an attempt to dilute out the endobacterial population. This technique was successfully used to dilute out the *Candidatus* Glomeribacter gigasporarum facultative endobacteria in the AMF host *Gigaspora margarita* (Lumini et al 2007). Due to the length of *in vitro* life cycle of *R. clarus* (3 to 4 months), serial single spore sub-culturing has been maintained and is currently in the 5th generation. MRE have been detected in every generation using PCR.

Infecting Endobacteria-Free AMF With MRE

The AMF species, *Rhizophagus irregularis* does not harbour any endobacteria. This species was used for infection, with the MRE harvested from the AMF species *Racocetra verrucosa*, and *Funneliformis mosseae*. Freshly sub-cultured *R. irregularis* was used for the infection, using plates with and without cellophane. The freshly inoculated *R. irregularis* was monitored for a few days, until new hyphae had germinated. The hyphae were cut to create openings, and the cut regions were filled with the filtrates of 50-100 spores, containing MRE. Each plate received either MRE populations from the two separate AMF species, or a mixture of the two MRE populations. Newly formed spores were checked for MRE presence using PCR, but none were ever detected.

A.2 - DISCUSSION

A functional study on the role of MRE in AMF would provide valuable information on the role of MRE in the fungi. Furthermore, any potential differences seen between AMF that harbour and do not harbour the bacteria can be used to determine whether the bacteria are parasites or

mutualists of their AMF hosts. In order to properly assess the above, we attempted to create isogenic AMF lines that harboured MRE (MRE+) or were free of MRE (MRE−). This was done by either curing AMF that were already MRE+, or infecting MRE− AMF lines with MRE.

Neither the curing of MRE+ lines of AMF nor infecting the MRE− lines of AMF were successful, despite the fact that these techniques were used in other systems where hosts harboured endobacteria (Lumini et al 2007, Partida-Martinez et al 2007). This suggests that, for the species of AMF that harbour MRE, this association may be obligate, and the partners are dependent on each other. This is further confirmed by our MRE metagenomic data (Chapter 3) that showed evidence of co-evolution of the two partners. Chapter 4 also provided evidence that the MRE-AMF association is ancient, and thus their co-evolution and dependence on each other would be expected. The inability of MRE− lines to be infected is likely due to both host and bacterial factors. AMF that are free of endobacteria have likely evolved defensive genes that inhibit endobacterial colonization, or lack gene products that are required for bacterial uptake and maintenance. Furthermore, as the MRE metagenomic data from Chapter 3 suggest, each MRE population has evolved for their specific AMF host, indicating that the MRE populations used for the infection study may have gene products that allow for AMF colonization, but only for a subset of AMF species. We believe that curing MRE+ lines of AMF is likely not possible, as the MRE-AMF are likely an obligate association. However, infection of MRE− AMF species may be possible, if MRE populations used are harvested from AMF species that are closely related to the MRE− AMF target species.

A.3 - MATERIALS & METHODS

AMF Materials

Rhizophagus clarus NB112A were obtained from the International Culture Collection of Vesicular Arbuscular Mycorrhizal Fungi (INVAM). An *in vitro* culture of *R. clarus* NB112A was started from single spores, and maintained as either multi-spore inoculum sub-cultures or single-spore inoculum sub-cultures. *Racocetra verrucosa* VA103A and *Funneliformis mosseae* MD209 were obtained from INVAM. These species are not cultivable *in vitro*, and were stored in the original soil samples at 4 ºC until use. *Rhizophagus irregularis* DAOM 181602 were obtained from Dr. David Douds and maintained *in vitro*.

AMF spore extraction from soil samples

AMF spores were extracted from dried soil samples by wet sieving and sucrose centrifugation (den Bakker et al 2010). 50 g of soil was shaken vigorously for 15 minutes in a flask containing 200 mL of distilled water. The soil suspension was poured through a 200 µm sieve, followed by a 90 µm sieve. The materials caught on the 90 µm sieves were centrifuged in tubes containing water and kaolin (Sigma-Aldrich) at 4400 rpm for 5 min. The pellet was resuspended in 2 M sucrose, and centrifuged at 4400 rpm for a further 3 min. The supernatant, containing the AMF spores, was placed on a 0.45 µm filter, and stored at 4 ºC until further processing.

AMF spore surface sterilization

Soil extracted AMF spores were surface sterilized using an antibiotic decontamination procedure (Bécard and Fortin 1988, Chabot et al 1992). First, the spores were pulse vortexed for 20 sec in 0.05% Tween® 20 solution (Sigma-Aldrich) to remove soil debris from the spores. Spores were

then treated in 2% chloramine-T (Sigma-Aldrich) for 10 min, followed by 3 treatments of 10 min exposures to an antibiotic mix (2% streptomycin sulphate (Sigma-Aldrich) and 1% gentamicin sulphate (Sigma-Aldrich)), with gentle shaking. The chloramine-T and antibiotic treatments were repeated for an additional 2 times. The spores were left in antibiotic treatment at 4 ºC for a further 24 h. Decontaminated spores were then placed in PBS solution (pH 7.0) and stored at 4 ºC until use.

AMF *in vitro* cultivation

Surface sterilized *R. clarus* and *R. intraradices* spores were maintained as *in vitro* cultures. Spores were grown on root-inducing T-DNA-transformed chicory roots, and co-cultivated on MSR medium (Cranenbrouck et al 2005), and incubated at 28 ºC.

MRE detection via polymerase chain reaction (PCR)

Single AMF spores were isolated from the *in vitro* cultures and manually crushed to expose MRE cells. Illustra™ GenomiPhi V2 Amplification Kit (GE Healthcare) was used, following the manufacturer's instructions, for whole genome amplification on the crushed material. MRE-specific 16S rRNA primers 109F1 5'- ACGGGTGAGTAATRCTTATCT -3', 109F2 5'- ACGAGTGAGTAATGCTTATCT -3', 1184R1 5'- GACGACCAGACGTCATCCTY -3', 1184R2 5'- GACGACCAAACTTGATCCTC -3', and 1184R3 5'- GATGATCAGACGTCATCCTC -3', were used to detect MRE genomic DNA in the pre-amplified material. JumpStart™ REDTaq® ReadyMix™ Reaction Mix (Sigma) was used following the manufacturer's directions, with an initial denaturation time of 3 min, a melting temperature of 60 ºC, and an extension time of 1 min. The cycle was repeated 35 times. The

primers were added to a final concentration of 0.75 uM for 109F1 and 1184R1, and 0.375 uM for 109F2, 1184R2, and 1184R3.

Isolation of MRE for infection

Spores were extracted from soil samples of *R. verrucosa* and *F. mosseae* and surface sterilized. The spores were then manually crushed in PBS (pH 7.0) and filtered through a 2 μm filter to remove cellular debris. Each filtrate used for infection contained MRE from 50-100 spores.

REFERENCES

Bécard G, Fortin JA (1988). Early events of vesicular-arbuscular mycorrhiza formation on Ri T-DNA transformed roots. *New Phytol* **108:** 211-218.

Chabot S, Bécard G, Piché Y (1992). Life cycle of *Glomus intraradix* in root organ culture. *Mycologia* **84:** 315-321.

Cranenbrouck S, Voets L, Bivort C, Renard L, Strullu D, Declerck S (2005). Methodologies for *in vitro* cultivation of arbuscular mycorrhizal fungi with root organs. In: Declerck S, Fortin JA, Strullu D (eds). *In Vitro Culture of Mycorrhizas*. Springer Berlin Heidelberg. pp 341-375.

den Bakker HC, Vankuren NW, Morton JB, Pawlowska TE (2010). Clonality and recombination in the life history of an asexual arbuscular mycorrhizal fungus. *Mol Biol Evol* **27:** 2474-2486.

Lumini E, Bianciotto V, Jargeat P, Novero M, Salvioli A, Faccio A *et al* (2007). Presymbiotic growth and sporal morphology are affected in the arbuscular mycorrhizal fungus Gigaspora margarita cured of its endobacteria. *Cell Microbiol* **9:** 1716-1729.

Partida-Martinez LP, Monajembashi S, Greulich KO, Hertweck C (2007). Endosymbiont-dependent host reproduction maintains bacterial-fungal mutualism. *Curr Biol* **17:** 773-777.